

# L'évaluation est-elle le moteur de l'échec scolaire ?

Dachet Dylan  
Journée en école du CECP – 30 mai 2023



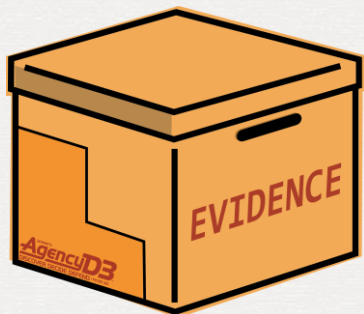
EQUALE\_

# Programme de la journée

- Le Paradigme de l'*Evidence-Based Education*
  - Qu'est-ce ?
  - Des programmes éducatifs étudiés en FW-B
  - L'analyse des croyances éducatives // un exemple
- Peut-on être juste et équitable en matière d'évaluation ?
  - Les biais de l'évaluation // rappel de la conférence
  - Les biais de l'évaluation sont-ils déterminés par des biais cognitifs généraux ?
- Un modèle global de l'évaluation
  - Focus sur l'erreur et son statut
  - Focus sur le renforcement positif
  - Focus sur les confusions entre les fonctions
  - (Focus sur les modes d'interaction)
  - (Focus sur les objectifs évalués)
- Ce que l'ont sait des méthodologies d'évaluation efficaces



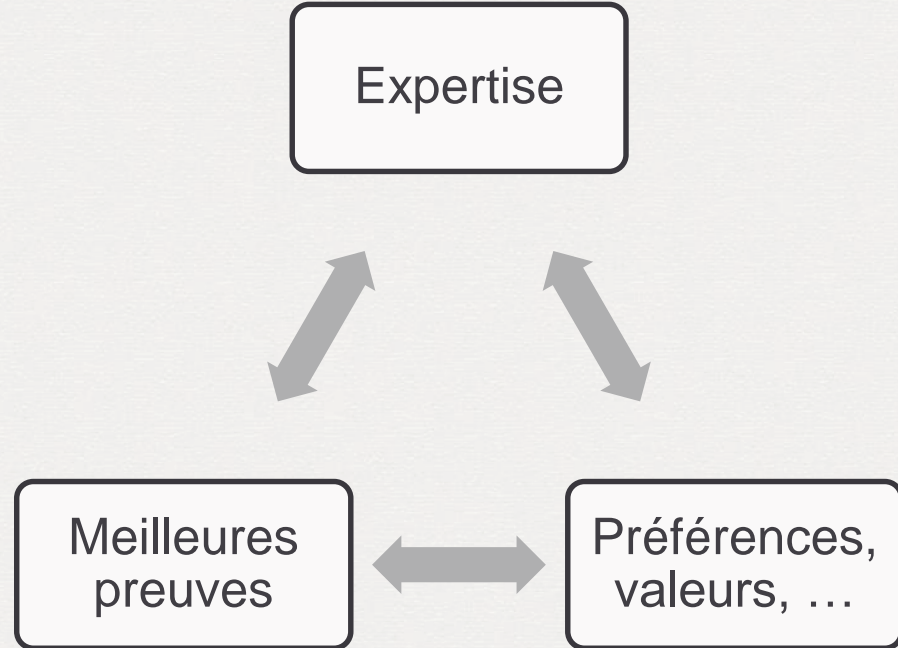
# Le Paradigme de l'*Evidence-Based Education*



# L' Evidence-Based Medicine

Médecine  
Fondée sur  
les Preuves

« l'utilisation **conscientieuse, explicite et judicieuse** des **meilleures preuves** actuelles pour **prendre des décisions** concernant les soins à apporter à des patients »

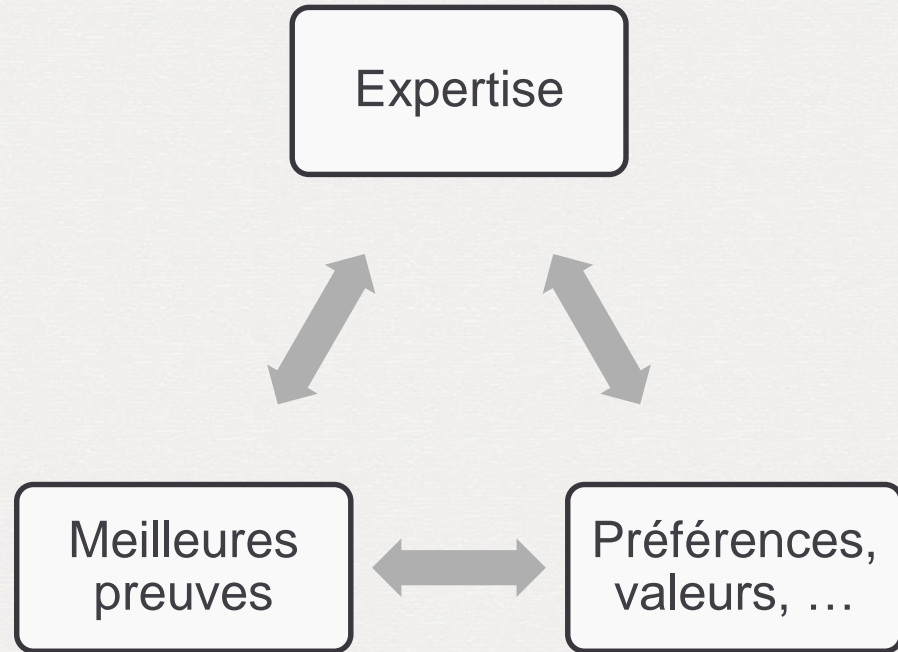


# L' Evidence-Based Education

Éducation  
Fondée sur  
les Preuves

« approche des **pratiques** et des **politiques éducatives** qui préconise que celles-ci doivent être **fondées sur les données les plus probantes produites par la recherche** »

Saussez & Lessard (2009, p. 111)



# L'Evidence-Based Education

Issu de Slavin (2015)

## Processus de recherche spécifique en 6 étapes :



**1. Revue de la recherche** antérieure quant à l'efficacité d'interventions éducatives spécifiques et répondant à un besoin clairement identifié.

Priorisation des domaines d'intervention.

Établissement de standards de qualité permettant la sélection des études.

Mise en place d'un centre d'expertise chargé de passer en revue les études disponibles.

**2. Communiquer** largement les résultats disponibles sur les dispositifs efficaces.

**3. Développer** les programmes d'intervention (adaptation socio-culturelle & critères d'évaluation).

Étude pilote de faisabilité à petite échelle.

Étude rigoureuse à grande échelle.

**4. Diffuser** les programmes d'intervention.

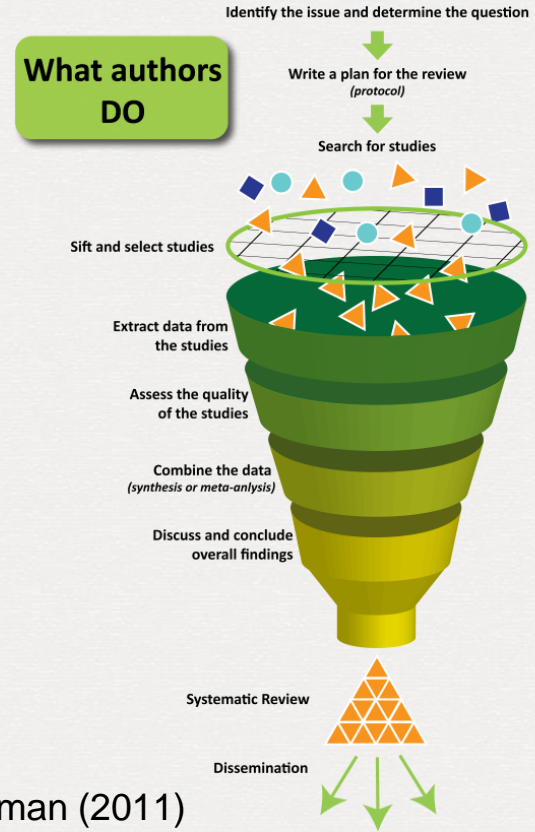
**5. Étendre le domaine** d'intervention.

**6. Évaluer l'impact** de la réforme **au niveau national**.

# La Revue Systématique de la Littérature

**Synthèse** (qualitative/quantitative)  
**exhaustive** des résultats de  
l'ensemble des études menées quant  
à une question/un sujet précis(e) et  
sélectionnées sur base d'un  
protocole strict et explicite.

(Li Wan Po, 1998 ; Maison, 2010 ; The Cochrane Collaboration, 2017)



Issu de Kaufman (2011)

# L'Evidence-Based Education

Issu de Slavin (2015)

## Processus de recherche spécifique en 6 étapes :

**1. Revue de la recherche** antérieure quant à l'efficacité d'interventions éducatives spécifiques et répondant à un besoin clairement identifié.

Priorisation des domaines d'intervention.

Définition de critères de qualité permettant la sélection des études.

Mise en place d'un centre d'expertise chargé de passer en revue les études disponibles.

**2. Communiquer** largement les informations disponibles sur les di

**3. Développer** les programmes d'intervention (// adaptés à la situation).

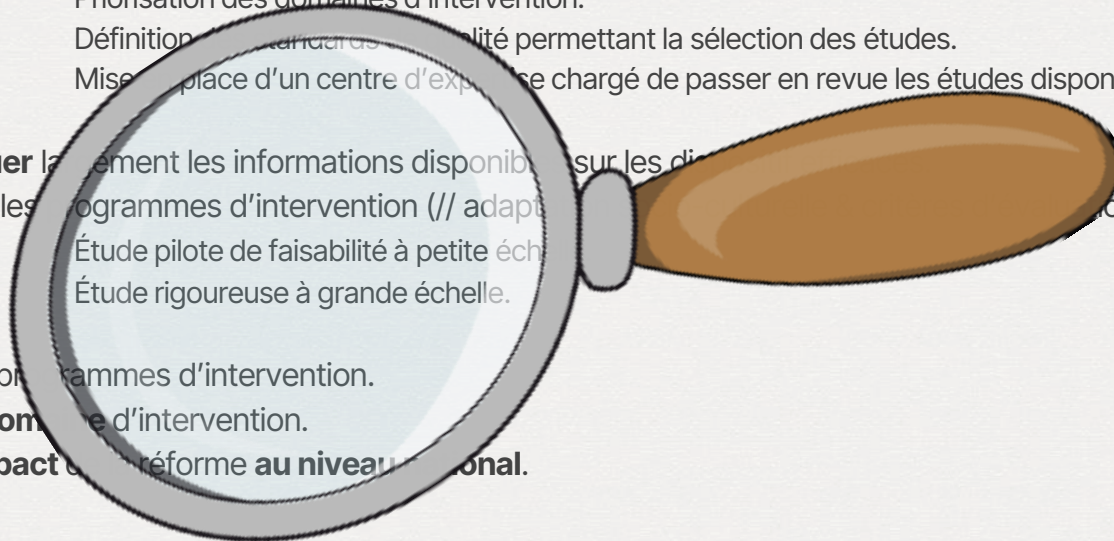
Étude pilote de faisabilité à petite échelle

Étude rigoureuse à grande échelle.

**4. Diffuser** les programmes d'intervention.

**5. Étendre le domaine** d'intervention.

**6. Évaluer l'impact** de la réforme **au niveau national**.



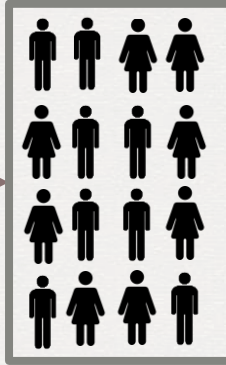
# Comment établir des liens de causalité ?

Mise en œuvre de dispositifs inspirés de la pédagogie Montessori

Année 0



Année 1

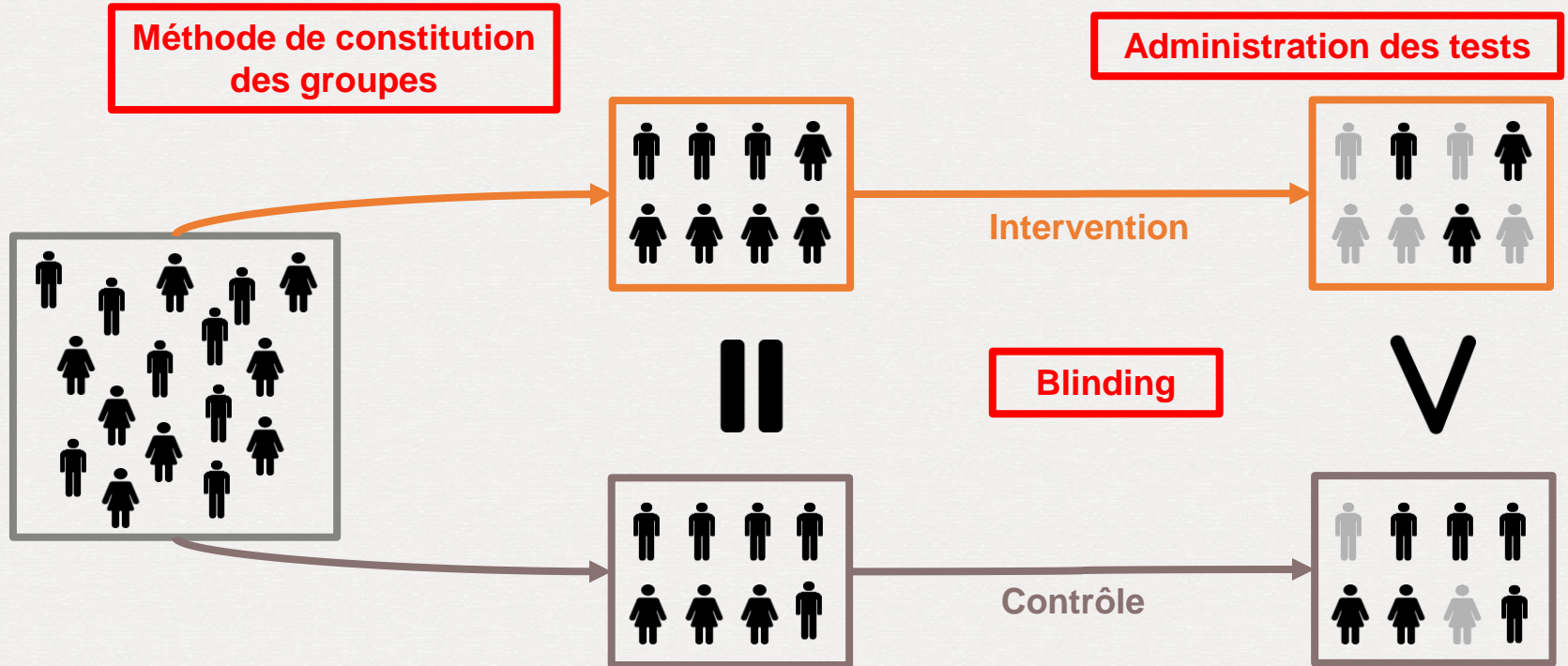


Année 2



# Comment établir des liens de causalité ?

Mise en œuvre de dispositifs inspirés de la pédagogie Montessori



# Comment établir des liens de causalité ?

Quels sont les effets des pédagogies d'inspiration Montessori ?

**Auteurs** : Courtier et al. (2021)

**Intervention** : dispositifs d'inspiration Montessori (a) une organisation thématique et ouverte de la classe ; (b) des classes d'âges mixtes ; (c) du matériel pédagogique, souvent multi-sensoriel et autocorrectif, spécifique

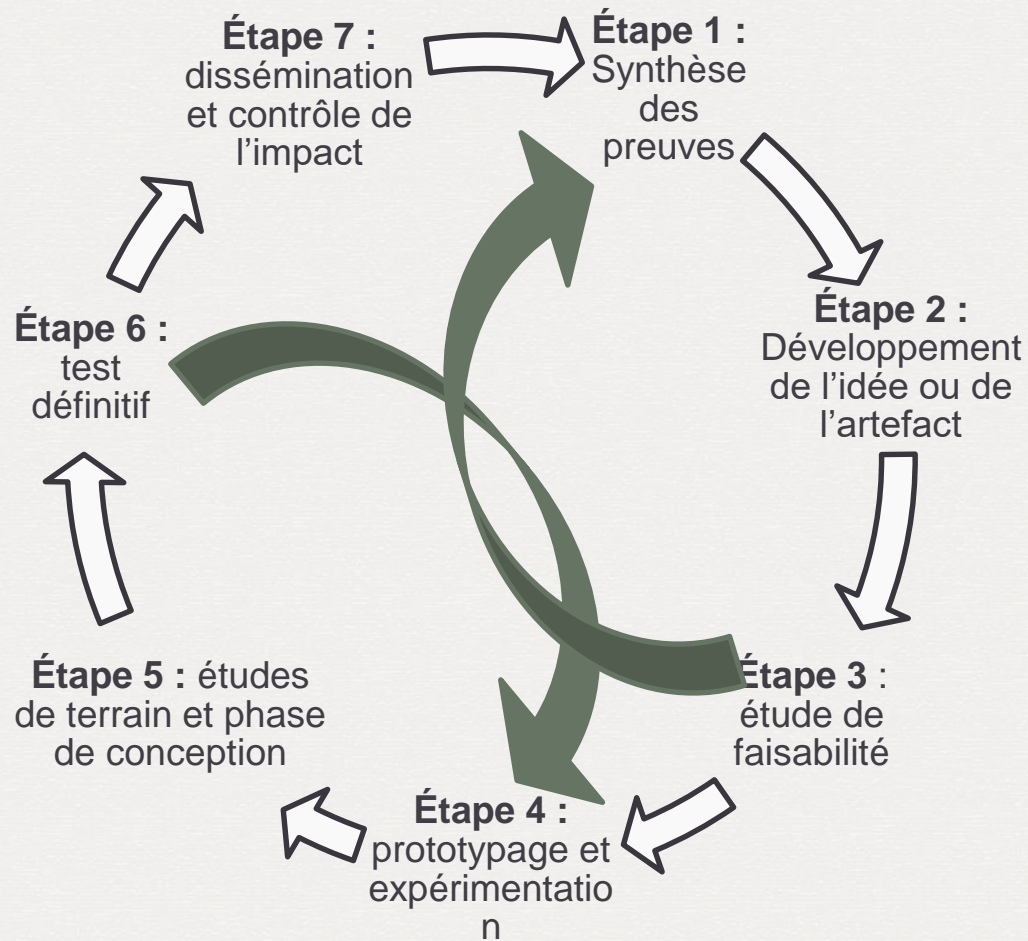
**Localisation** : Lyon

**Echantillon** : trois cohortes (2017, 2018, 2019) d'élèves → 176 élèves de 5 à 6 ans

**Constitution des groupes** : une école maternelle privée et trois classes d'une école publique bénéficiant de l'intervention, 5/6 classes d'une école publique servant de contrefactuel

**Résultats** (70 élèves – 33 GE vs. 37 GC):

- Aucun effet positif sur les variables académiques (langage, conscience phonologique, mathématiques-résolution de problèmes ...) sauf en lecture ( $d = 0.68$ )
- Aucun effet positif sur les variables cognitives (mémoire à court-terme, autorégulation ...)
- Aucun effet positif sur les variables sociales (partage, justice, compétences sociales ...)



Issu de Gorard & Taylor (2004) et de Cook & Gorard (2007)

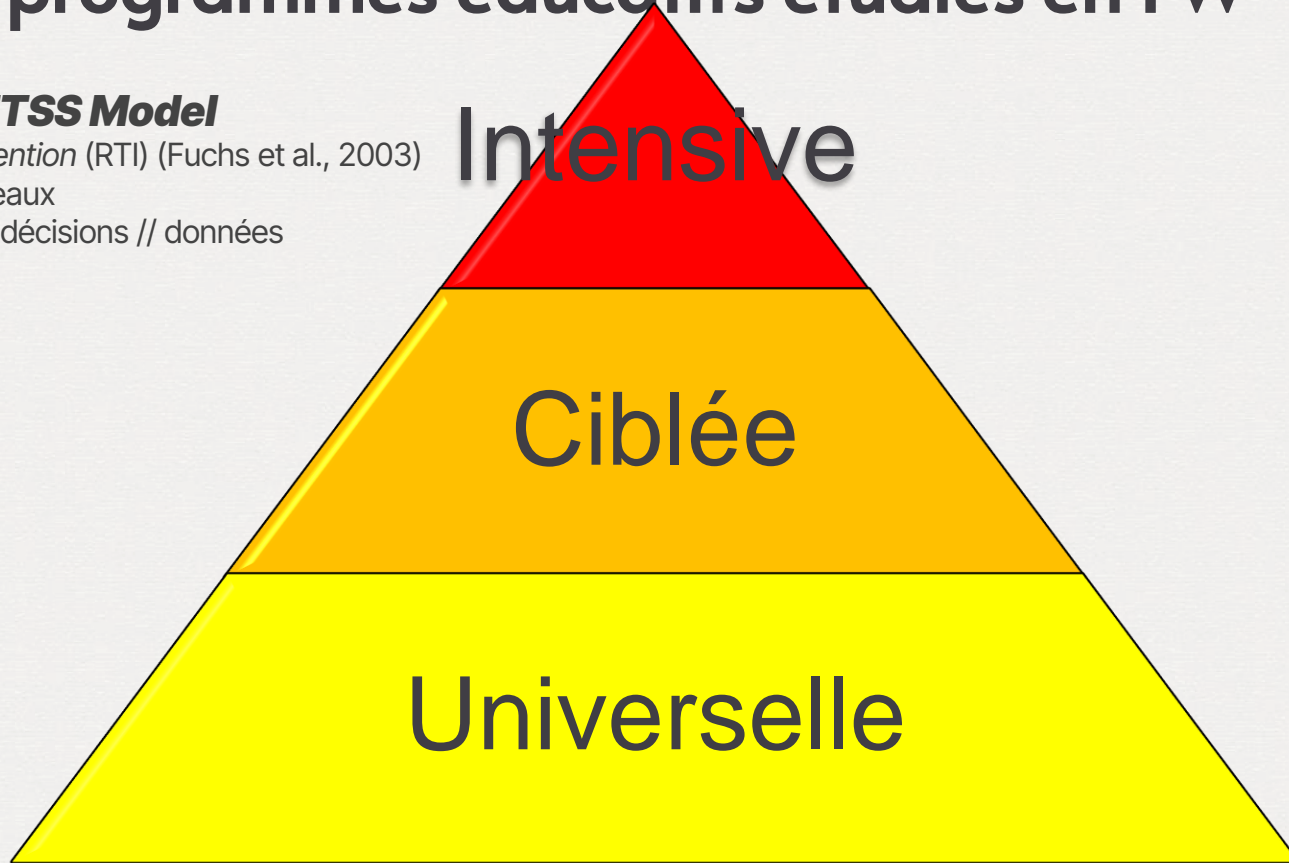
# Des programmes éducatifs étudiés en FW-B

## **MTSS Model**

*Respons-To-Intervention* (RTI) (Fuchs et al., 2003)

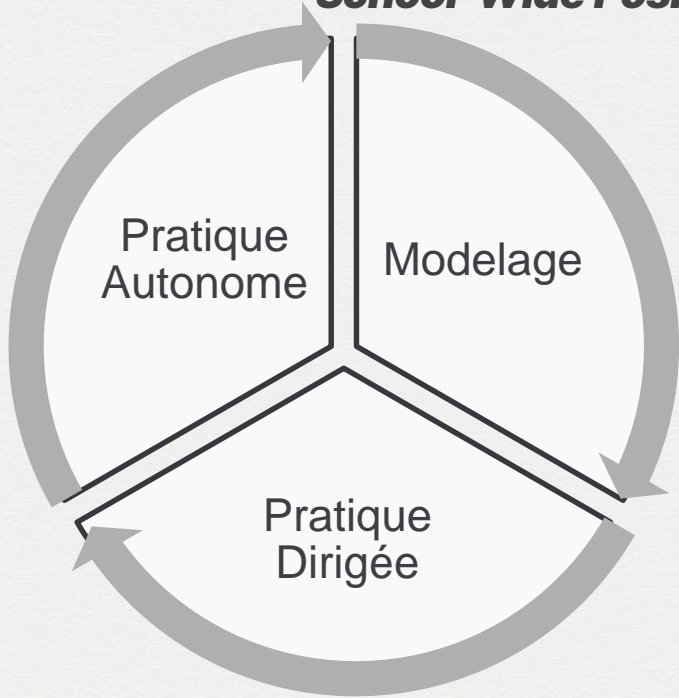
Modèle en trois niveaux

Modèle de prise de décisions // données






# Des programmes éducatifs étudiés en FW-B

## *School-Wide Positive Behaviour Intervention and Support*



(Deltour et al., 2021 ; Gauthier et al., 2013 ;  
Newcomb Park Primary School, 2019)

Newcomb Park Primary School				
School Wide Positive Behaviour Matrix				
	IN THE CLASSROOM AND DURING SPECIALIST	IN THE YARD	IN TRANSITION, AT THE OFFICE AND BEFORE OR AFTER SCHOOL	AS AN AUDIENCE
<b>RESPECT</b> 	We listen to and follow instructions We allow others to learn We accept different students learning needs We speak nicely We keep our hands, feet and objects to ourselves We respect personal space and privacy	We include others We share equipment and spaces We keep our hands, feet and objects to ourselves We look after school property, equipment and grounds We speak nicely We return equipment inside	We use inside voices We wait for our turn at the office and use our manners We look after school property, equipment and grounds We listen to and follow instructions We use the bathrooms appropriately	We sit appropriately as an audience and listen We clap and respond appropriately at the right times We use our manners We stand still, keep our hands by our side and face the front while singing the National Anthem
<b>RESPONSIBILITY</b> 	We arrive on time, organised and ready to learn We join in and try our best We take care of our own and others property We move safely around the classroom/space We ask before leaving the classroom/specialist	We report dangerous behaviour to a teacher We stop playing when the music starts We put rubbish in the bin We play appropriately in the right areas We wear our school hat in terms 1 and 4	We move safely through all areas of the school We make sure we are in the right place at the right time We only leave the classroom with teacher permission We walk our bikes and scooters within the school grounds	We follow the teacher's instructions at all times We choose a good place to sit to avoid distractions
<b>RESILIENCE</b> 	We persist with all tasks, demonstrating a growth mindset and challenging ourselves We work and play cooperatively by taking turns and playing fairly, letting others join in and following the rules We are positive bystanders and tell a teacher if we need help We think of another solutions We walk away if we are feeling upset or angry and get help from an adult We use the cool down nook/ calm space according to classroom guidelines			

# Votre/vos question(s)

Comment rendre efficaces les dispositifs d'aide aux élèves en difficulté ?  
Approfondir l'enseignement explicite

# Des programmes éducatifs étudiés en FW-B

## **PARLER**

Parler Apprendre Réfléchir Lire Ensemble pour Réussir

Enseignement explicite et systématique :

- Des habiletés liées à l'acquisition du code écrit (conscience phonologique, principe alphabétique et fluidité de lecture en contexte).
- Des habiletés liées à la compréhension pratiquée à l'oral (vocabulaire, entraînement explicite à l'appropriation de stratégies de compréhension, lectures partagées).

Activités en petits groupes

Temps d'entraînement individualisé // besoins spécifiques

Séances de renforcement (2/semaine) au sein de petits groupes de 4/5 élèves en difficulté

# Des programmes éducatifs étudiés en FW-B

## *KiVa*

### Intervention

Repose sur les recherches quant à la position sociale des enfants agressifs et harceleurs et sur le rôle des parties-prenantes.

**Axe universel** : 20 heures de cours, discussions, travaux de groupe, jeux de rôle, courts-métrages et jeu vidéo – je sais, je peux, je fais.

**Axe centralisé** : discussions individuelles, suivis en petits groupes et soutien de pairs de haut niveau.

**Axe formation des professionnels.**

### Méthode

Quasi-expérimentale – GE = 4207 ; GC = 4030

1 an – prétest, posttest mi, posttest final

Self-reported and peer-reported bullying and victimization ; ...

Variable	AE	Variable	AE
<i>SR Victim</i>	- 0.12	<i>PR Defen</i>	+ 0.08
<i>SR Bully</i>	- 0.05	<i>Anti-Bull</i>	+ 0.07
<i>PR Victim</i>	- 0.15	<i>Empathy</i>	+ 0.05
<i>PR Bully</i>	- 0.12	<i>Self-effi</i>	+ 0.07
<i>PR Assist</i>	- 0.11	<i>Well-being</i>	+ 0.11
<i>PR Reinfor</i>	- 0.13		

# L'analyse des croyances éducatives

- Les styles d'apprentissage
  - Les intelligences multiples
  - L'hypnopédie
  - Le *MindMapping*
  - La pyramide de Maslow
  - L'effet Mozart
  - Les hommes sont plus performantes en mathématiques
  - Les femmes sont plus performantes en lecture
  - 10 000 heures de pratique pour devenir un expert
  - ...
- 

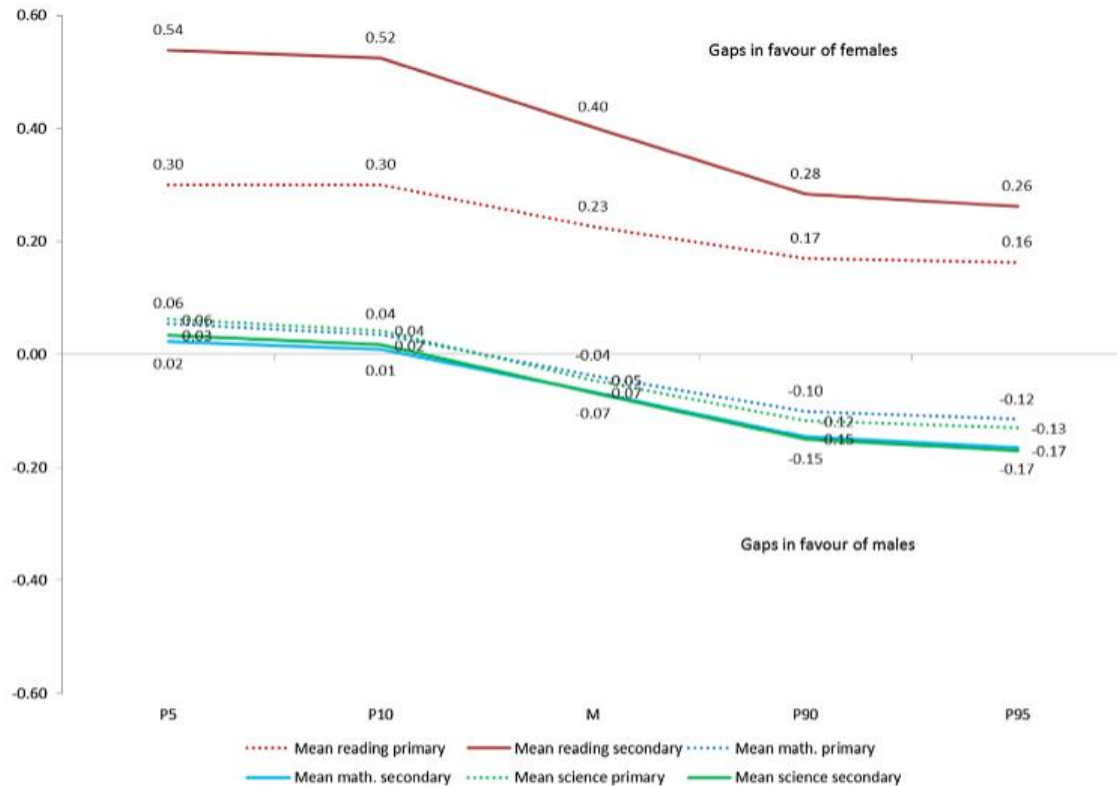
Industrie de la publication (Pashler et al., 2008)  
Décentration de la responsabilité  
Théorie de la relativité absolue  
Distinction entre préférence et efficacité

Catégorisation :

- Appartenance catégorielle
- Homogénéité intra-catégorielle
- Aucun recouvrement inter-catégoriel

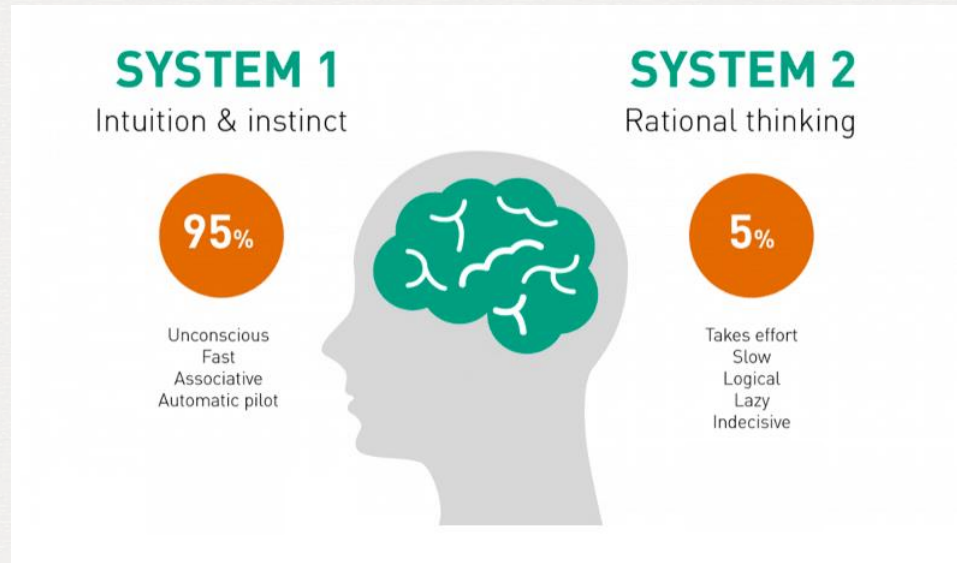
Instrument de mesure (Sander, 2018)  
71 styles définis (Coffield et al., 2004)

Absence de preuves de l'interaction attendue



**Fig. 1** Gender effect sizes in mean and extreme tails of the distribution in reading, math. and science. Gender effect sizes at percentiles 5, 10, 90, 95 and on the mean scores, by content area and education level. Positive values indicate higher scores for females; negative values indicate higher scores for males. *Data IEA (TIMSS/ PIRLS) and OECD (PISA) surveys, 1995–2012*

# Peut-on être *juste* et *équitable* en matière d'évaluation ?



Kahneman (2012)

# Les biais de l'évaluation // Rappel

Equipe A



Monde idéal

Equipe B



Biais

Le biais et le bruit – l'écart  
systématique et la  
dispersion aléatoire – sont  
deux composantes  
différentes de l'erreur

(Kahneman et al., 2021, p. 9-11)

Equipe C



Bruit

Equipe D



Biais & Bruit

# Les biais de l'évaluation sont-ils déterminés par des biais cognitifs généraux ?

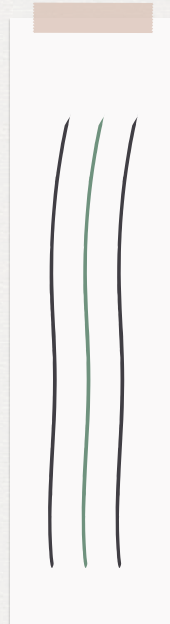
## Biais de confirmation

Tendance à privilégier les informations et/ou exemples qui confirment nos croyances au détriment de celles qui pourraient les remettre en question (Durand, 2019, 121-135)

### Exemples :

- Syndrome de l'étudiant en médecine
- Étude de Lord et al. (1979) sur la peine de mort
- Étude de Sunstein et al. (2017) sur le réchauffement climatique

**Explication** : psychologie évolutionniste (Boutang & De lara, 2009)



## Recherche en docimologie

### Effets de Halo

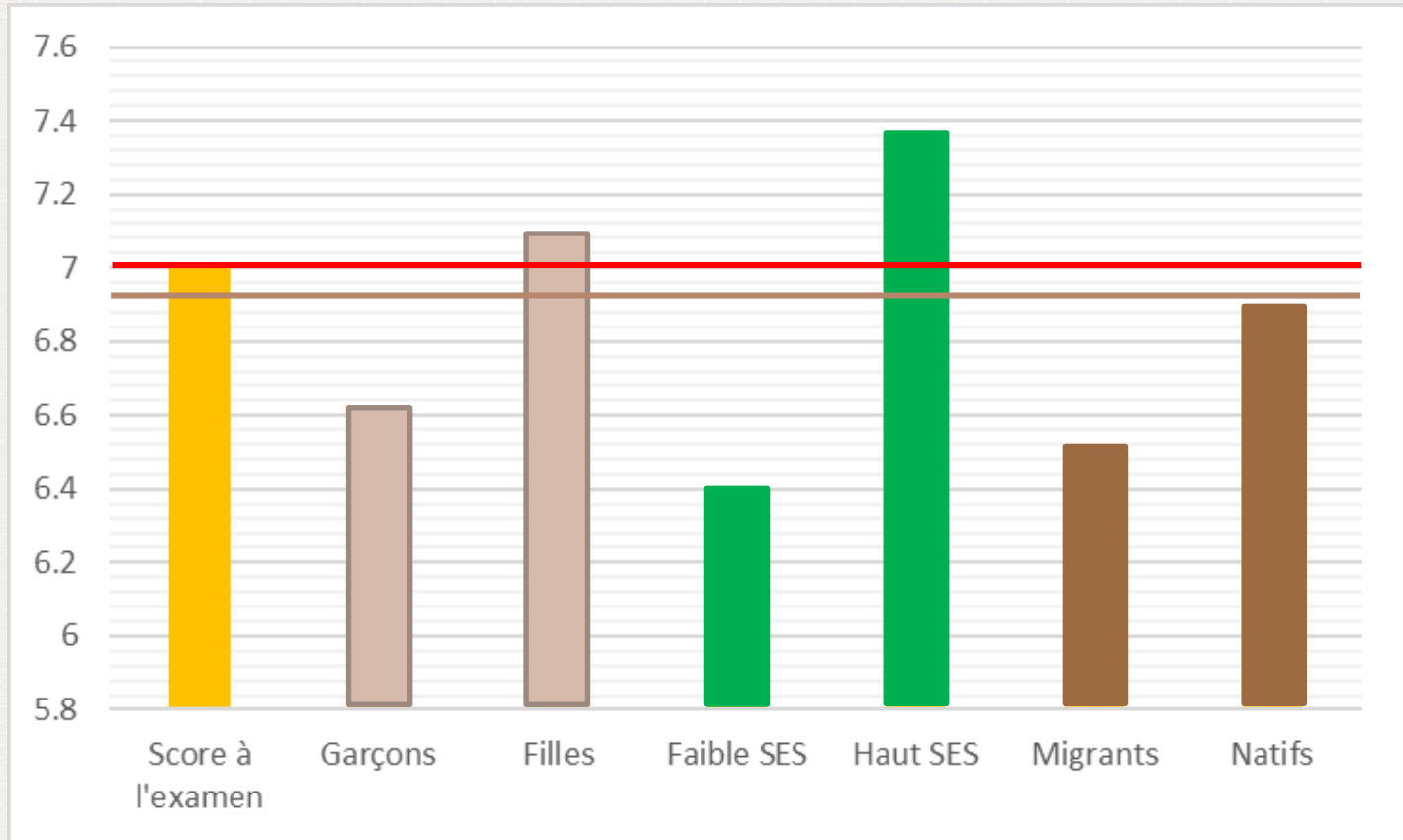
- Origine ethnique (Dutrévis, 2015 ; Sprietsma, 2013)
- SES (Autin et al., 2019 ; Rangvid, 2015)
- Niveau d'éducation parentale (Rangvid, 2015)
- Covid-19 (Baye & Dachet, 2022)
- Genre en mathématiques (Lafontaine & Monseur, 2009)

Les élèves porteurs d'un stéréotype négatif ont tendance à être sous-évalués comparativement à des élèves qui ne sont pas porteurs de ce stéréotype.



# Étude de Rangvid (2015)

Comparaison entre : (1) les données issues des examens étatiques externes et (2) les notes des enseignants



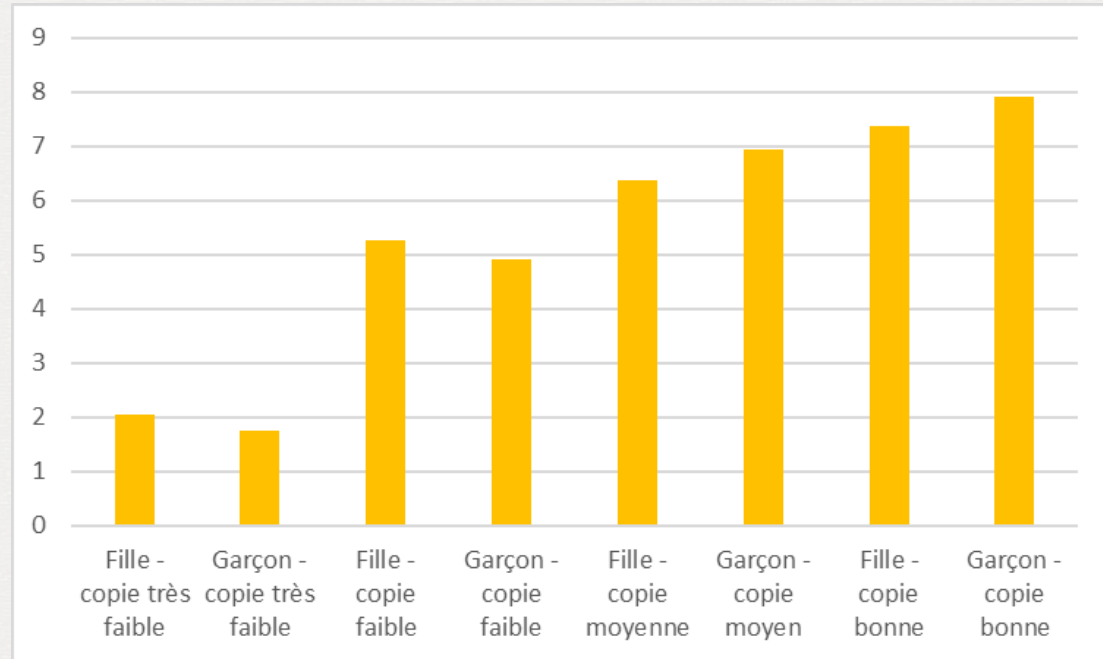
# Étude de Lafontaine et Monseur (2009)

## Etude expérimentale :

Copies fictives construites par un professeur de mathématiques : TF, F, M, B

Seule information disponible → prénom clairement masculin (Bernard et Nicolas) ou féminin (Chloé et Emilie)

48 professeurs de mathématiques



# Étude de Dachet et al. (2023)

Existe-t-il un biais de l'évaluation lié à l'origine ethnique des élèves indépendant de leur origine sociale et de leur niveau de performance ?

**Design :** RCT ; réplication d'une étude menée en 2022 et non publiée.

**Population :** étudiants en formation initiale, étudiants en masters en sciences de l'éducation et étudiants en psychologie

**Taille de l'échantillon :** 500 participants

**Instrument :** trois productions écrites d'élèves de 5<sup>e</sup> année de l'enseignement primaire (1500 copies)

**Grille d'évaluation :** issue des évaluations externes non certificatives de la 5P

Forte



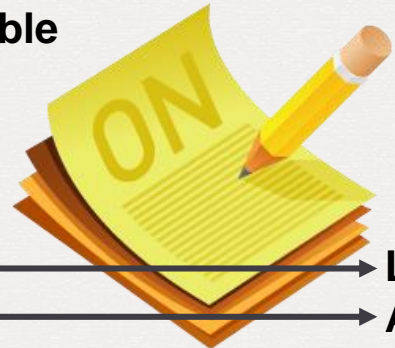
Ayoub  
Jordan  
Lucien

Moyenne



Jordan  
Lucien  
Ayoub

Faible



Lucien  
Ayoub  
Jordan

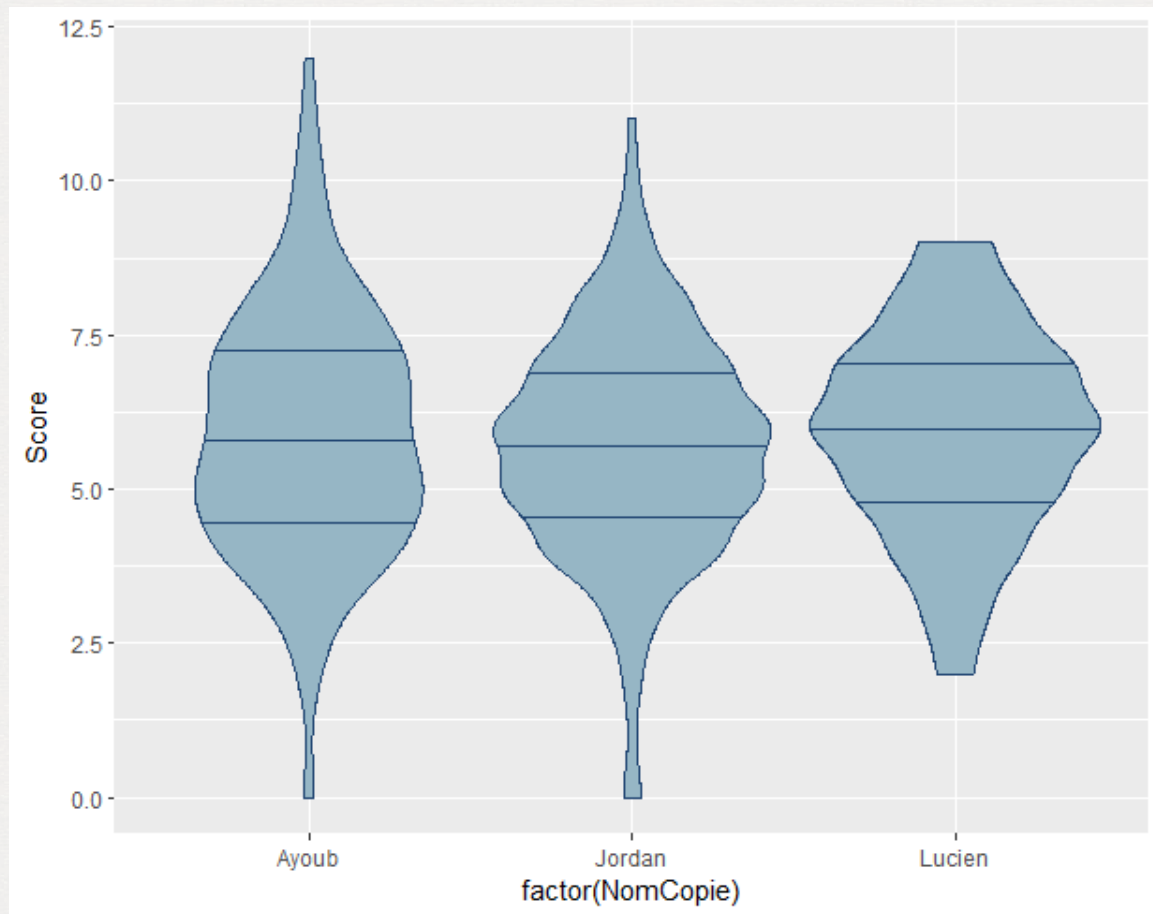
# Étude de Dachet et al. (2023)

Comparaison entre les 9 copies

	N°	Moyenne	ET
Ayoub Faible	1	5.884	2.022
Ayoub Moyen	2	10.625	1.823
Ayoub Fort	3	11.501	1.741
Jordan Faible	4	5.683	1.815
Jordan Moyen	5	9.851	1.703
Jordan Fort	6	11.853	2.092
Lucien Faible	7	5.92	1.678
Lucien Moyen	8	9.241	2.066
Lucien Fort	9	12.282	2.282

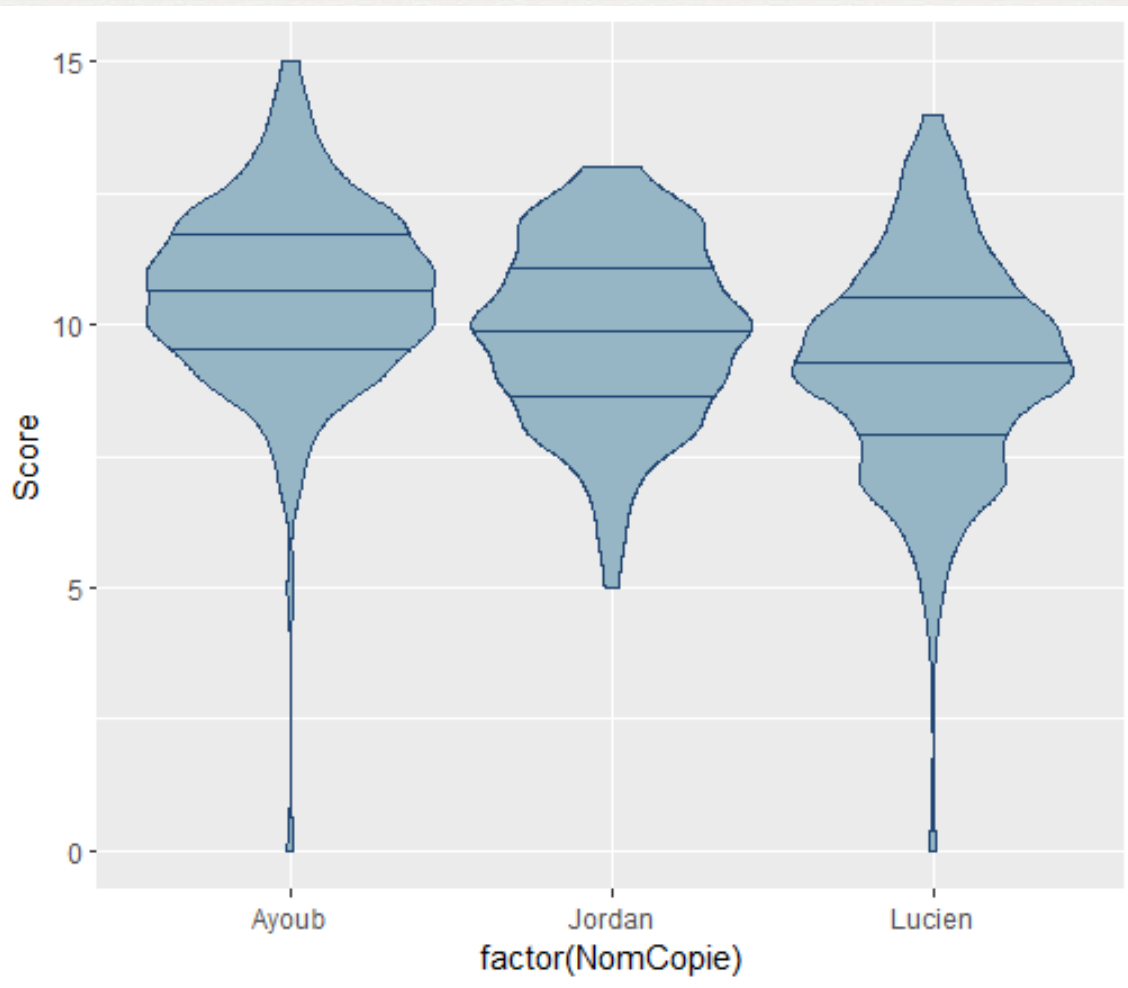
## Les copies faibles

Comp.	Hedges'g
<i>Lu vs. Jo</i>	+0.135
<i>Lu vs. Ay</i>	+0.020
<i>Jo vs. Ay</i>	-0.105



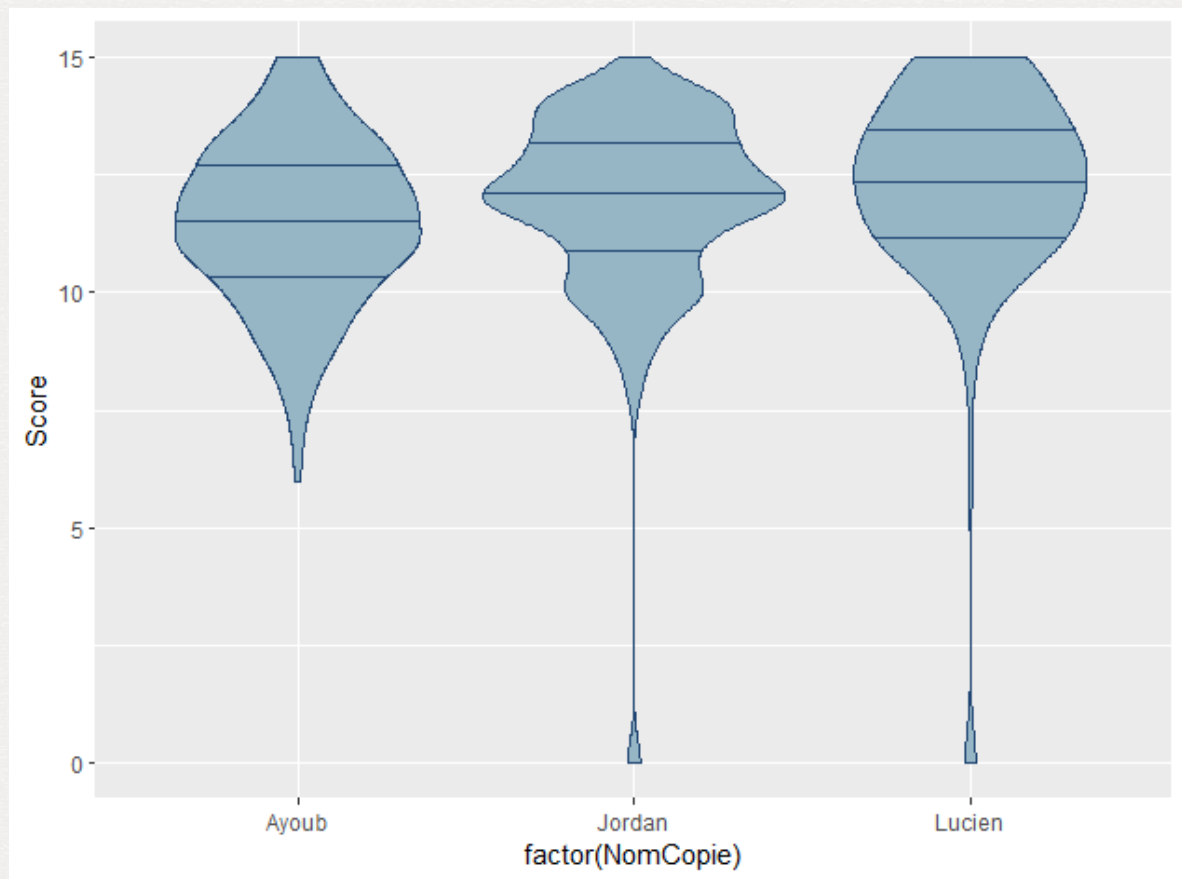
## Les copies moyennes

Comp.	Hedges'g
<i>Lu vs. Jo</i>	-0.323
<i>Lu vs. Ay</i>	-0.703
<i>Jo vs. Ay</i>	-0.433



## Les copies fortes

Comp.	Hedges'g
<i>Lu vs. Jo</i>	+0.195
<i>Lu vs. Ay</i>	+0.384
<i>Jo vs. Ay</i>	+0.183



# Les biais de l'évaluation sont-ils déterminés par des biais cognitifs généraux ?

- Il semble bien exister un effet de Halo lié à l'origine ethnique indépendamment de l'effet de Halo lié au SES.
- L'utilisation de grilles d'évaluation et de consignes de correction permettent de diminuer l'effet de Halo, mais n'ont pas suffi afin de le supprimer totalement.
- Biais de confirmation → les copies fortes et pour les copies moyennes.
- Quid des copies faibles ?
  - L'absence de différence entre les trois conditions pourrait être due à la grille utilisée → trop faible que pour amener débat
- Perspectives :
  - développer un continuum de copies faibles
  - répliquer les résultats d'Autin et ses collègues (2019)
- Quid des caractéristiques prises en compte en contexte réel ?
- Quelles autres solutions pédagogiques ?



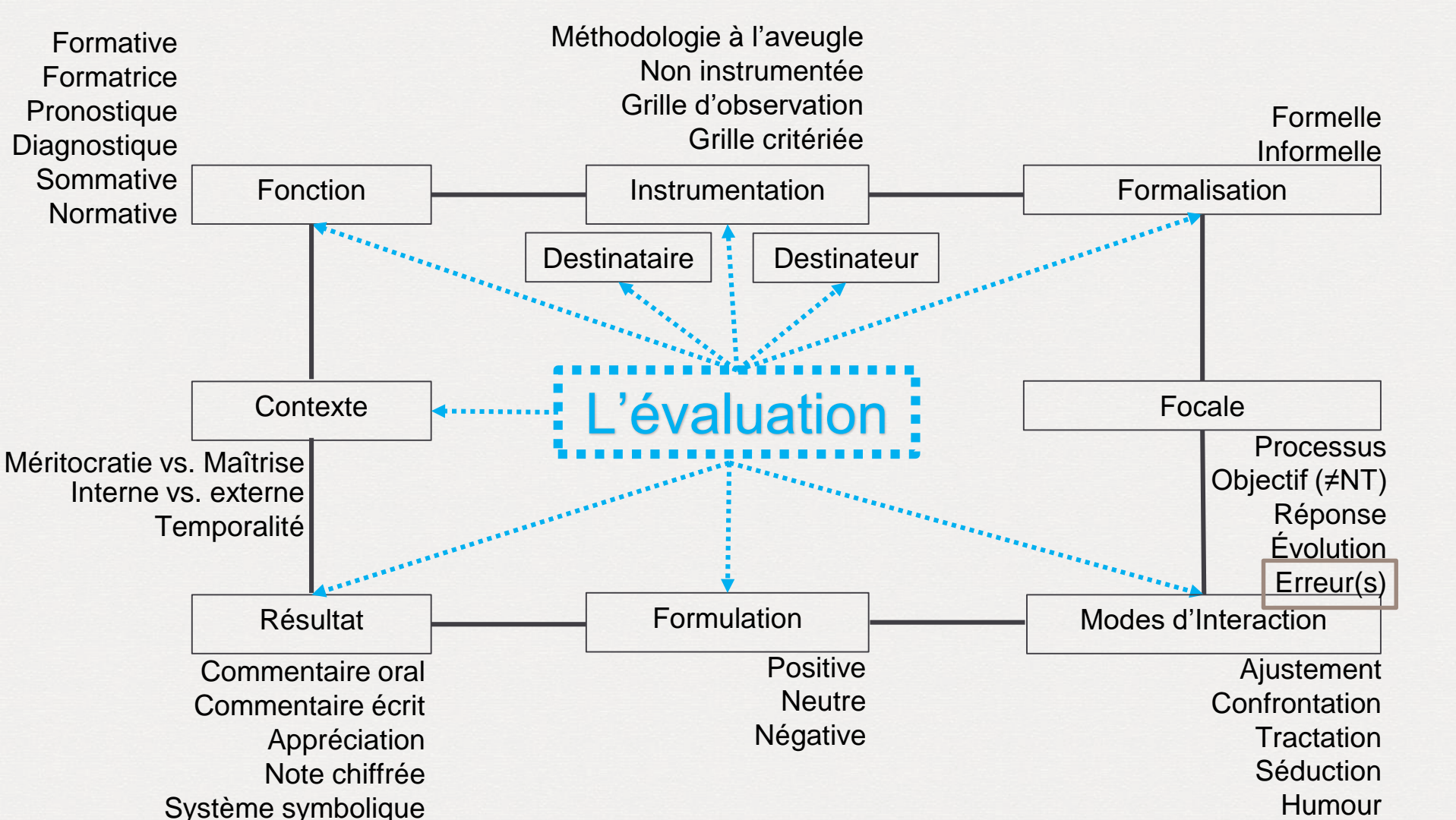
# Votre/vos question(s)

L'évaluation formative serait-elle au service de la lutte contre l'échec scolaire ?

Faut-il évaluer avec des points ? Ou pas ? Quid du stress de l'élève ?



# Un modèle global de l'évaluation



# Analyse des erreurs

Astolfi (2020)

- « Les erreurs commises ne sont plus des fautes condamnables, ni des bogues regrettables : elles deviennent des symptômes intéressants d'obstacles auxquels la pensée des élèves est affrontée » (Astolfi, 2020, p. 22).

	<b>La faute</b>	<b>Le bogue</b>	<b>L'obstacle</b>
<b>Statut</b>	L'erreur est déniée		L'erreur est positivée
<b>Origine</b>	Elève	Planification	Difficulté objective d'appropriation du contenu
<b>Traitement</b>	Évaluation <i>a posteriori</i> pour la sanctionner	Traitement <i>a priori</i> pour la prévenir	Travail <i>in situ</i> pour la traiter

Astolfi (2020, p. 32)

# Analyse des erreurs

Astolfi (2020)

Michel Fabre définit 6 caractéristiques des obstacles :

- Intériorité
- Facilité // repérage des régularités au lieu des règles
- Positivité
- Ambiguïté
- Polymorphie
- Récursivité

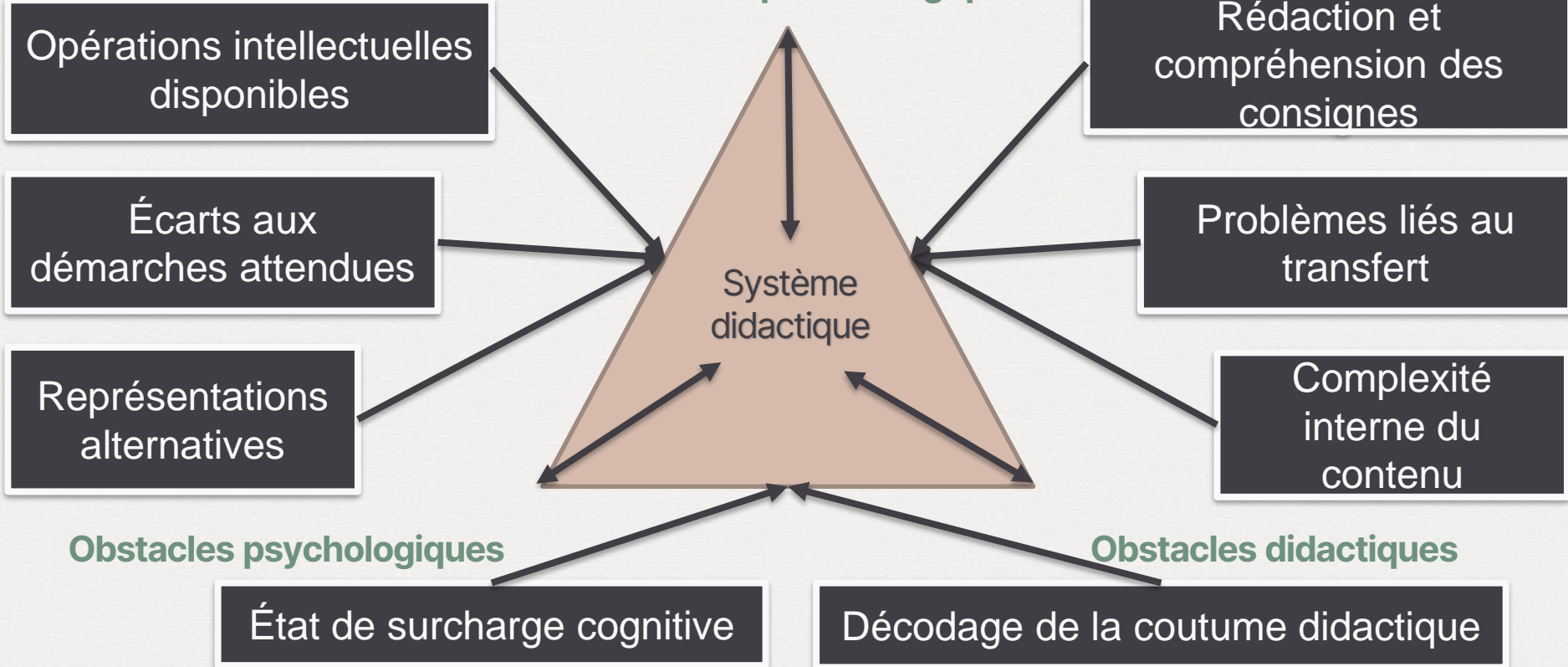


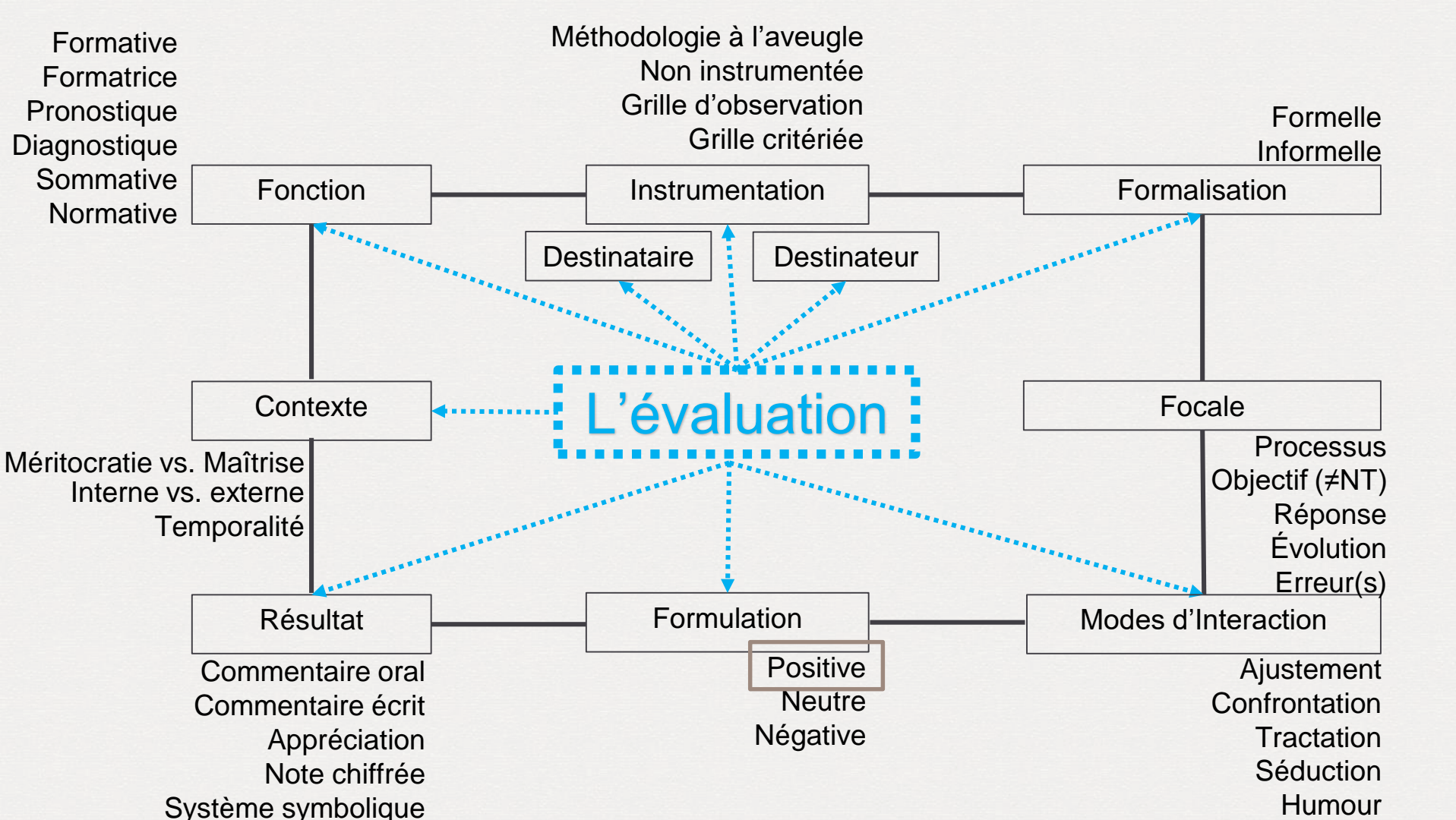
- Travail chronophage dont la mise en œuvre doit être réfléchie et anticipée
- Travail qui vise à formuler des hypothèses
  - Toute hypothèse mérite d'être confirmée par une récolte ultérieure d'informations

# Analyse des erreurs

Astolfi (2020)

Obstacles épistémologiques





# Le renforcement positif

Revue systématique (Royer et al., 2018)

**QR** : *Est-ce que la pratique du renforcement positif spécifique au comportement peut-être considérée comme fondée sur des preuves ?*

→ L'éloge = **subordonnée** au comportement, **spécifique**, **sincère**, **variée**, et **crédible**  
= **facteur malléable** plutôt facteur non contrôlable  
= manière dont le **comportement rencontre l'attente** ou **affecte la réussite**

**Méthode** : recherche dans 21 bases de données bibliographiques → 318 références

recherche à la main → 32 références

- 282 articles exclus pour des raisons de pertinence
- 29 articles exclus après lecture des synthèses
- 33 articles exclus après lecture complète

→ 6 articles rencontrent les critères d'inclusion (enseignants ; comportements ; K-12 ; recherche expérimentale)

# Le renforcement positif

## Revue systématique (Royer *et al.*, 2018)

Étude	Mesure	Effet
Madsen <i>et al.</i> (1968)	Comportements inappropriés	+
Van der Mars (1989)	Comportements hors de la tâche	+
Houghton <i>et al.</i> (1990)		+
Wheatley <i>et al.</i> (2009)	Comportements inappropriés	+
Haydon & Musti-Rao (2011)	Perturbations	+
Hollingshead <i>et al.</i> (2016)	Comportements dans la tâche	/

# Le renforcement positif

Revue systématique (Royer et al., 2018)

**Conclusion** : « *Les six études examinées montrent que le renforcement positif peut être mis en œuvre avec un minimum d'effort et pratiquement aucune perturbation des enseignements, alors qu'il entraîne des changements rapides dans les résultats des élèves. D'autres études plus rigoureuses sur le plan méthodologiques sont nécessaires pour déterminer si le renforcement positif comme utilisé dans les écoles traditionnelles de la maternelle à la fin de l'enseignement secondaire supérieur peuvent être considérés comme des pratiques evidence-based.* »

# Le renforcement positif

Enquête sur les adolescents (Fefer et al., 2016)

**QR** : *Quelles sont les préférences des élèves adolescents en matière de renforcements positifs et de récompenses relatifs au comportement scolaire ?*

**Méthode** : enquête administrée à 764 élèves de la 5<sup>e</sup> année (P5) à la 12<sup>e</sup> (S6) année d'un district scolaire américain

**Résultats** :

- Les adolescents **apprécient** tant les renforcements positifs que les récompenses
  - Les adolescents accordent, souvent, **plus d'importance à l'opinion des enseignants** concernant le travail scolaire qu'à l'opinion d'autres (pairs, parents, ...)
  - Les **renforcements positifs « silencieux »** (ou privés/discrets) sont plus appréciés que les renforcements positifs publics.
- **Ce dernier constat semble surtout vrai pour les adolescents, moins pour les plus jeunes (Burnett, 2001)**

# Le renforcement positif

## Caractéristiques

## Définition

(Understood for all, 2019)

### Sincère et véritable

Repose sur une forme de vérité // éviter les exagérations (Brummelman et al., 2014)

### Subordonné au comportement

Dépend intimement du comportement observé

### Spécifique

Décrit exactement le comportement observé et son intérêt

### Pas directif

Ne décrit pas des tâches futures à accomplir

### Centré sur le processus

Renforce la démarche, le progrès, ... plus que le résultat et/ou la comparaison avec les autres

### Immédiat

Est formulé immédiatement après l'observation du comportement

### Cohérent avec les attentes

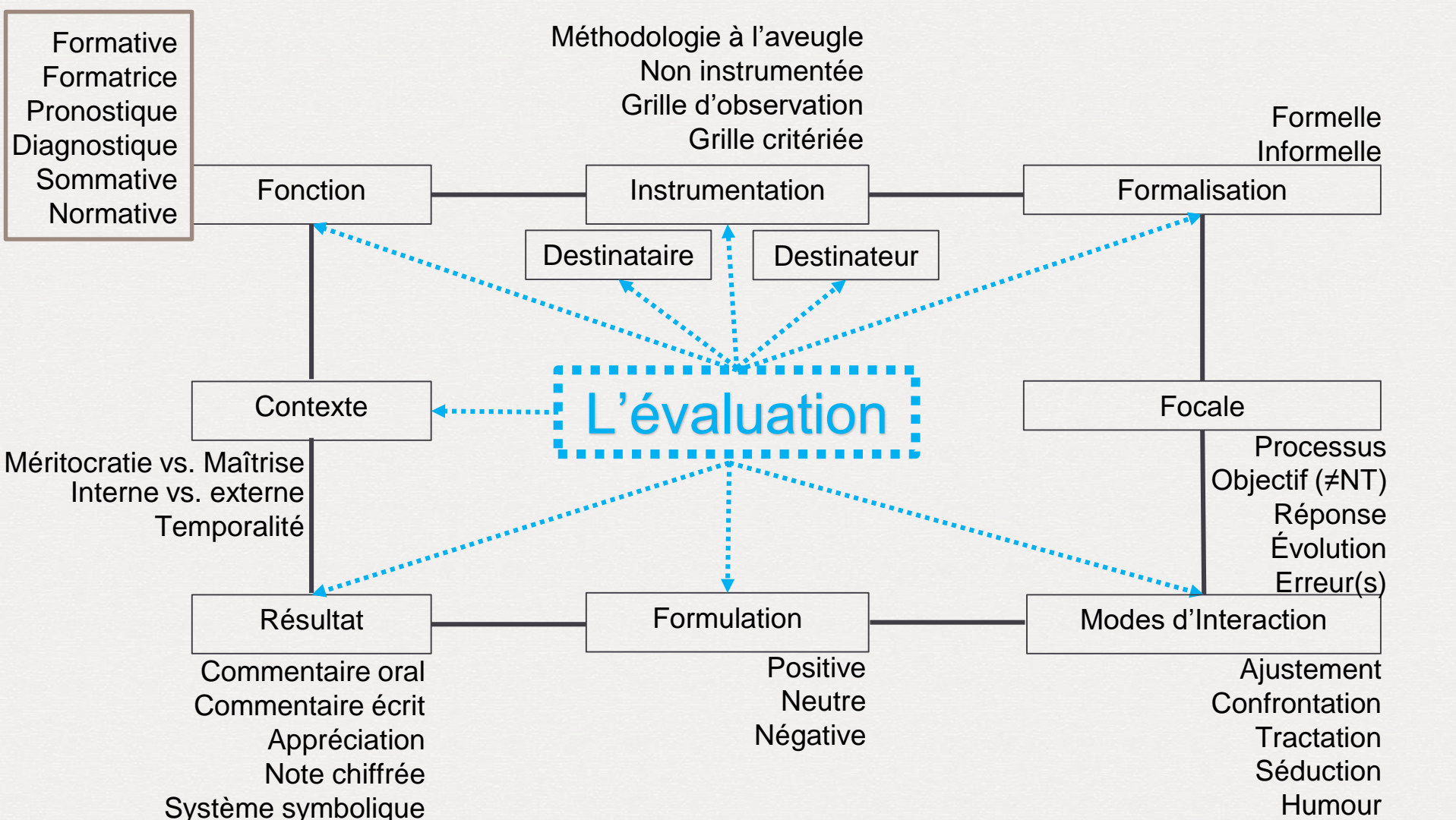
N'est utilisé que pour valoriser un comportement observé qui correspond au comportement attendu et explicité

### Inattendu

N'est pas prévisible

### Varié

Porte sur l'ensemble des comportements attendus



# Les fonction évaluatives

	Fonction	Confusion
<b>Normative</b>	Réguler les hiérarchies sociales, académiques, ...	VS. Sommative
<b>Sommative</b>	Réguler les parcours de formation par la vérification de l'atteinte des objectifs	VS. Pronostique/ Normative/Formative
<b>Pronostique</b>	Réguler les parcours de formation par la vérification a priori des compétences	VS. Sommative/ Diagnostic
<b>Diagnostic</b>	Réguler <i>a priori</i> l'enseignement et/ou l'apprentissage	VS. Pronostique
<b>Formative</b>	Réguler <i>a posteriori</i> l'enseignement et/ou l'apprentissage	VS. Sommative
<b>Formatrice</b>	Réguler <i>a posteriori</i> l'enseignement et/ou l'apprentissage et/ou le processus évaluatif	VS. Sommative

# Les fonction évaluatives

(Crahay, 2019)

Évaluation dans une perspective ...	
D'excellence	De maîtrise
<p>Évaluation normative</p> <ul style="list-style-type: none"><li>→ Classement des individus les uns/aux autres</li><li>→ Distribution gaussienne des notes</li><li>→ Questions discriminantes</li></ul>	<p>Évaluation <b>sommative</b> à référence <b>critériée</b></p> <ul style="list-style-type: none"><li>→ Permettre de situer chaque élève par rapport aux compétences assignées comme objectifs</li></ul>
<p>Évaluation pronostique</p> <ul style="list-style-type: none"><li>→ Apte à suivre ?</li><li>→ Sélection.</li></ul>	<p>Évaluation <b>formative</b></p> <ul style="list-style-type: none"><li>→ Comprendre les difficultés éprouvées par l'élève<ul style="list-style-type: none"><li>• Diagnostic</li><li>• Analyse des réponses</li><li>• Interactions avec les élèves</li></ul></li><li>→ Réorienter l'apprentissage et/ou l'enseignement</li></ul>

# Les fonction évaluatives

Conception initiale de Bloom versus conception élargie de l'évaluation formative dans la littérature scientifique francophone (Allal & Mottier Lopez, 2005)

Conception initiale de Bloom	Conception élargie
Insertion de l'EF <u>après</u> la phase d'enseignement	Intégration de l'EF durant <u>tout</u> l'apprentissage
Utilisation de <u>tests</u> formatifs	Utilisation de <u>divers moyens</u> de recueil d'informations
Feed-Back + correction → <u>remédiation</u>	Feed-back + adaptation de l'enseignement → <u>régulation</u>
Gestion de l'EF par l' <u>enseignant</u>	Participation active des <u>élèves</u> à l'EF
Maîtrise des objectifs par <u>tous les élèves</u>	<u>Différenciation</u> de l'enseignement et des objectifs
<u>Remédiation bénéfique</u> aux élèves qui ont été évalués	<u>Régulation à 2 niveaux</u> : pour les élèves évalués et pour les futurs élèves

# Les fonction évaluatives

(Crahay, 2019)

## Évaluation formative formelle

Conçue pour produire des traces et indices sur l'apprentissage de l'élève  
→ *épreuves papier-crayon débouchant sur un feed-back critérié explicite*

### Recueillir

Des informations sur les élèves dans un temps planifié

### Interpréter (analyser)

Les informations recueillies (souvent hors de la présence des élèves)

### Agir

en planifiant une action visant à soutenir l'élève dans l'atteinte des buts d'apprentissage

## Évaluation formative informelle

Traces et indices générées pendant l'activité quotidienne de classe  
→ Pleinement intégrées aux processus d'enseignement et d'apprentissage en contexte de classe

### Solliciter

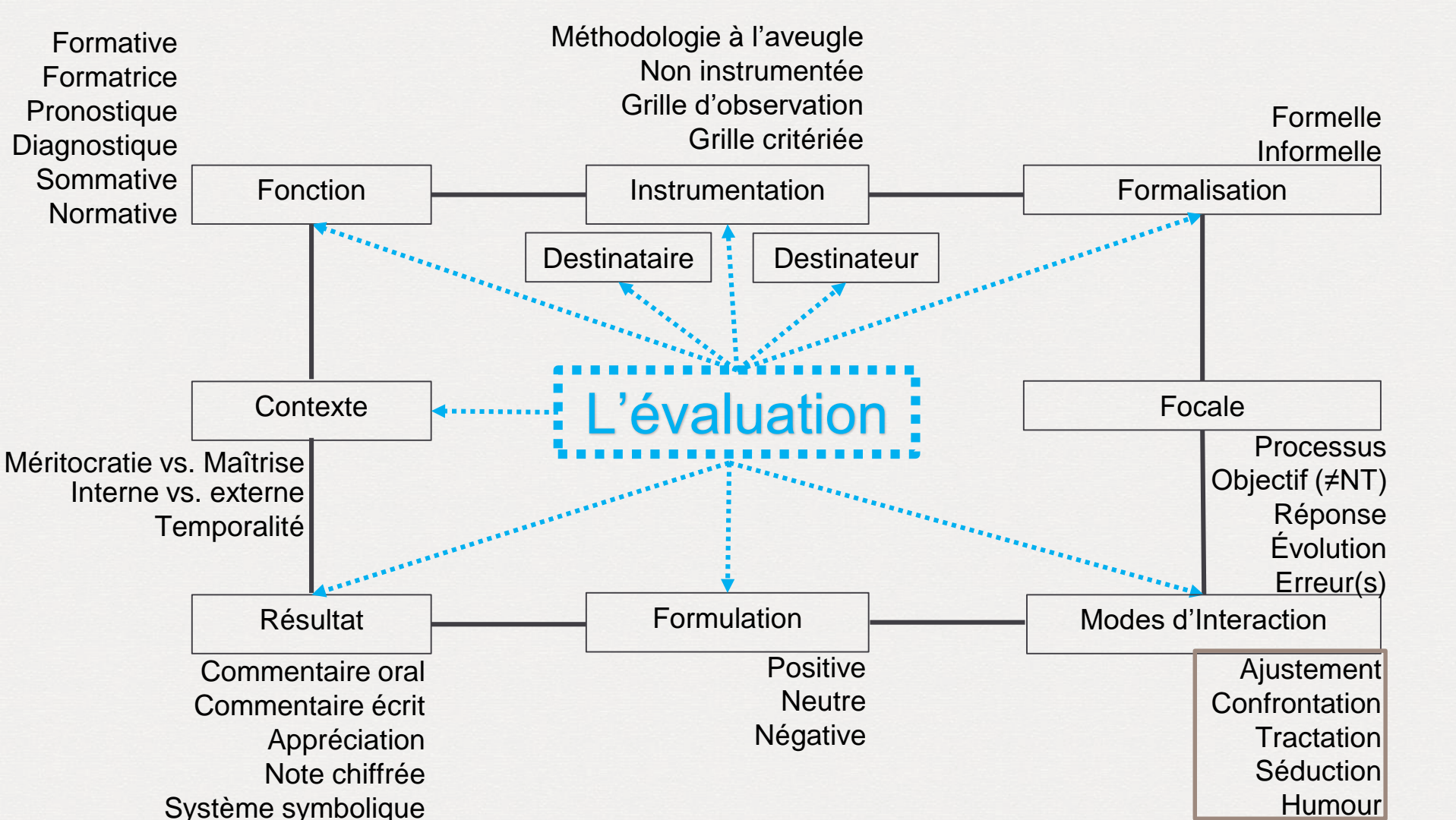
Des réponses verbales des élèves afin de récolter de l'information

### Reconnaitre

Les réponses des élèves au regard des concepts enseignés

### Utiliser

Immédiatement l'information dans le cours des activités continues de la classe



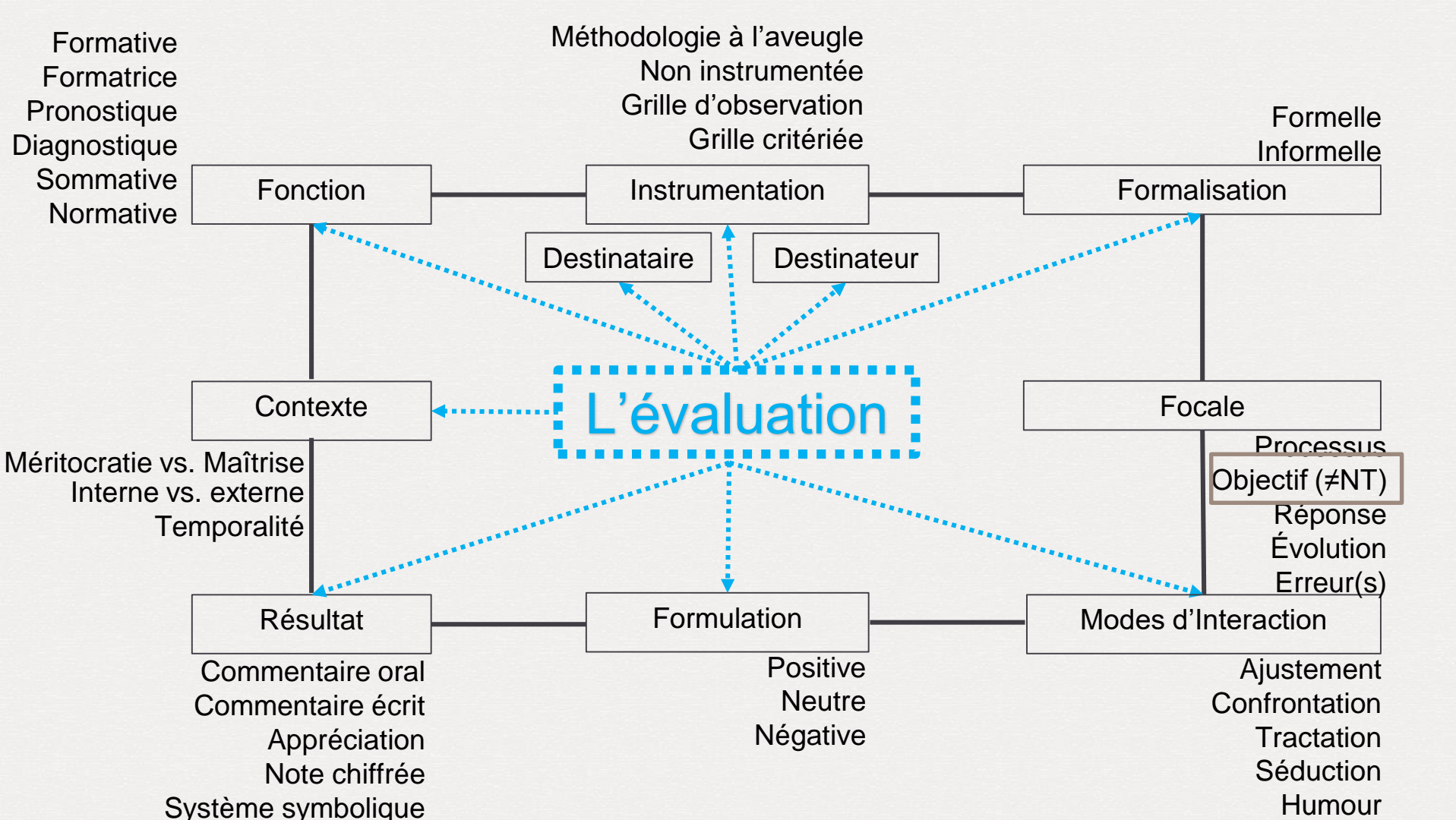
# Modes d'interactions

<b>Objectif</b>	Appréhender les modes d'interaction « enseignant-élèves » lors de l'évaluation formative non-instrumentée
<b>Participants</b>	10 enseignants
<b>Design</b>	Étude de cas (qualitative)
<b>Instrument(s)</b>	3 bandes vidéo réalisées dans chaque classe Identification par les enseignants des épisodes d'évaluation formative non instrumentée Entretiens individuels d'explicitation (de l'action et de leur compréhension de la dynamique avec les élèves)
<b>Analyse</b>	Microsociologie de Goffman (1973) : identification des modes d'interaction au travers de la définition de la situation par chaque enseignante Codage des données par une approche inductive : ajustements successifs, jusqu'à saturation (Laperrière, 1997), de la grille d'analyse au fur et à mesure de la confrontation avec un nouveau verbatim.

Issu de Morrissette (2013)

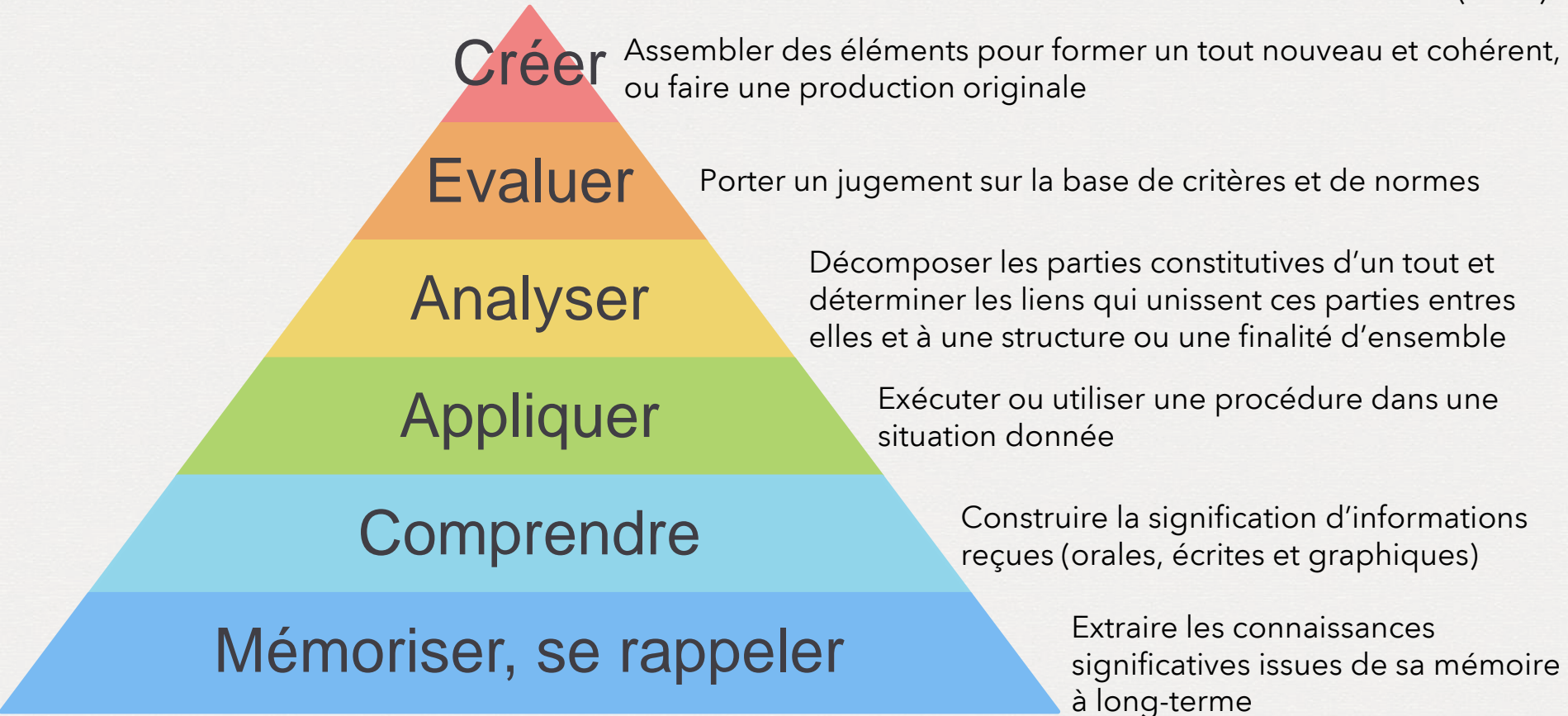
# Modes d'interactions

<b>Mode</b>	<b>Définition</b>
<b>Ajustement</b>	Enseignant et élève mobilisent leurs positions respectives dans un mouvement de recherche de sens compatible
<b>Confrontation</b>	Enseignant et élève mobilisent leurs positions respectives dans un mouvement d'opposition
<b>Tractation</b>	Enseignant et élève mobilisent leurs positions dans un mouvement de compromis, de conciliation
<b>Séduction</b>	Jeux socioaffectifs auxquels se livrent l'enseignant (valorisation de l'élève, ...) et l'élève (agit conformément aux attentes, ...) les détournant, pour un moment, de la visée de soutien aux apprentissages de l'évaluation formative.
<b>Humour</b>	Enseignant et élève jouent de leurs positions respectives pour plaisanter, taquiner, voire créer une situation absurde qui fait rire.



# Taxonomie de Bloom

Anderson & Krathwohl (2001)



# Votre/vos question(s)

Comment évaluer les enfants « qui auraient dû recommencer une année » lorsqu'ils n'ont pas les compétences pour l'année suivante ?



# Les méthodologies d'évaluation **efficaces**

# Évaluation formative ?

Revue systématique (Lee et al., 2021)

**QR :** (1) Quel est l'effet de l'évaluation formative sur l'apprentissage des élèves ? ; (2) Quelles sont les caractéristiques de l'évaluation formative et comment elles influencent de façon différentielle l'efficacité du processus d'évaluation formative ?

**Méthode :** recherche dans 5 bases de données bibliographiques + Google Scholar → 3730 références  
recherche dans les revues antérieures : 57 références  
sélection de 25 références qui rencontrent les CI (Exp. ; Mesure d'apprentissage ; K-12)

## **Résultats :**

126 ampleurs de l'effet ont été extraites des 25 études incluses. L'effet moyen calculé est de 0.29 ( $p < 0.001$ ). En d'autres mots, comparés aux étudiants qui n'ont pas participé à l'évaluation formative, les étudiants qui y ont participé ont démontré une augmentation de 29% d'écart-type de leurs performances aux tests d'apprentissage.

Table 3. Mean effect size estimation.

Estimation	# of ESs	# of Studies	# of Projects	Mean ES	SE
Overall	126	33	25	.29***	(.05)
<i>Literacy</i>	46	12	9	.33***	(.08)
<i>Mathematics</i>	50	14	12	.34***	(.09)
<i>Science</i>	18	4	3	.13***	(.00)
<i>Arts</i>	12	3	1	.29***	(.08)

\*\*\*  $p < .001$

Table 4. Analyses of moderator variables (Meta-regression).

Moderator variables		Predicted mean	SE	Difference
Source of FA feedback	(1) Teacher FA practice	0.18*	(0.09)	(2) > (1) ≈ (3)
	(2) Student-initiated self-assessment	0.61***	(0.16)	
	(3) Mixed	0.13	(0.08)	
Formality of FA evidence	(1) Formal FA evidence	0.40**	(0.13)	No difference
	(2) Informal FA evidence	0.18	(0.20)	
	(3) Mixed	0.26***	(0.05)	
Cycle length of FA feedback	(1) Medium-cycle length	0.52***	(0.15)	No difference
	(2) Short-cycle length	0.24***	(0.04)	
	(3) Not defined	0.24	(0.19)	
Professional development	(1) No PD	0.18	(0.11)	No difference
	(2) One-time PD	0.27***	(0.07)	
	(3) On-going supports	0.30***	(0.07)	
Computer	(1) Paper-based	0.29***	(0.06)	No difference
	(2) Computer-based	0.21*	(0.10)	
Instructional adjustment	(1) No adjustment	0.16	(0.18)	No difference
	(2) Planned adjustment	0.35*	(0.16)	
	(3) Unplanned adjustment	0.15	(0.13)	
	(4) Mixed	0.29***	(0.06)	
School	(1) K-6	0.29***	(0.05)	No difference
	(2) 7-12	0.23***	(0.06)	
Classroom	(1) Regular classroom	0.25***	(0.04)	No difference
	(2) Special education classroom	0.35**	(0.12)	

Values in light gray and dark gray are small and medium effect sizes, respectively; this only applies to statistically significant values.

\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ .

# Évaluation formative ?

Revue systématique (Karaman, 2021)

**QR :** (1) Quel est l'effet de l'évaluation formative sur l'apprentissage des élèves en Turquie ? ; (2) Quelles caractéristiques sont susceptibles de modérer l'efficacité de l'évaluation formative ?

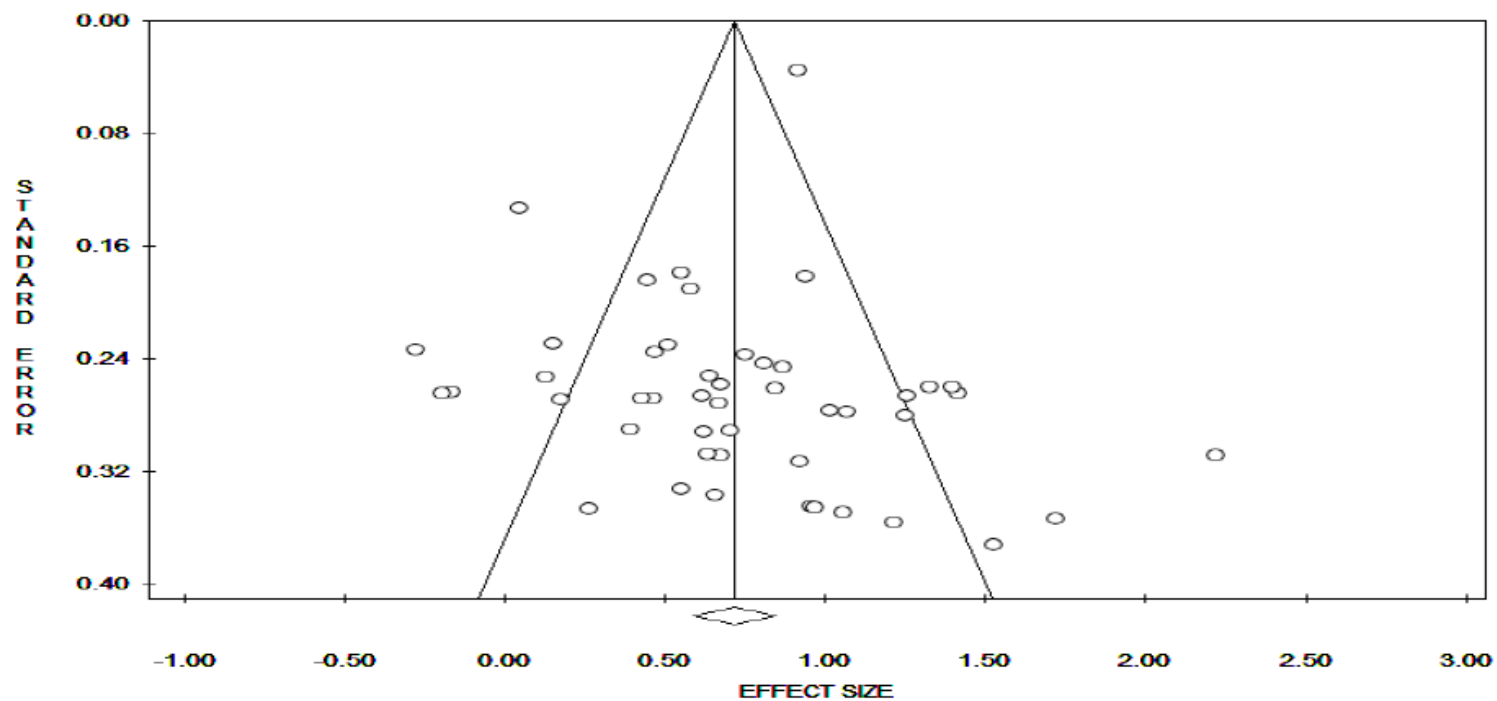
**Méthode :** recherche dans 8 bases de données bibliographiques + Google Scholar → 105 références  
recherche dans les revues antérieures : X références  
sélection de 32 références qui rencontrent les CI (RCT/QED ; Mesure d'apprentissage ; ES)

## Résultats :

47 ampleurs de l'effet ont été extraites des 32 études incluses. L'effet moyen calculé est de 0.72 (SE = 0.07 ;  $p < 0.05$ ). En d'autres mots, comparés aux étudiants qui n'ont pas participé à l'évaluation formative, les étudiants qui y ont participé ont démontré une augmentation de 72% d'écart-type de leurs performances aux tests d'apprentissage.

L'hétérogénéité des effets observés ( $Q = 188.91$ ) est significative ( $df = 46, p < 0.001$ ).

**Figure 2.** *A funnel plot indicating standard error and observed effect size.*



**Table 3.** Overall effect sizes and heterogeneity results related to the effectiveness of formative assessment practices.

Model	k	Mean ES	SE	Lower Limit	Upper Limit	p	Heterogeneity			
							Q value	df	p	I <sup>2</sup>
Fixed	47	.79	.03	.74	.84	.00	188.91	46	.00	75.65
Random	47	.72	.07	.59	.85	.00				

\*p<.05; k= number of effects; ES= effect size

**Table 4.** Results of the subgroup analysis for features of formative assessment interventions.

	k	Mean ES	SE	Lower Limit	Upper Limit	p	Q value	df	p
Adult initiated feedback	20	.69	.09	.50	.87	.000			
Student initiated feedback	10	1.16	.17	.83	1.49	.000			
Computer initiated feedback	12	.42	.14	.14	.71	.003			
Mixed feedback	5	.83	.19	.46	1.21	.000			
Heterogeneity test							11.54	3	.009

\*p < .05; k= number of effects; ES= effect size; SE= standard error

**Table 5.** Results of the subgroup analysis for education level.

Education Level	k	Mean ES	SE	%95 Confidence Interval		p	Heterogeneity		
				Lower Limit	Upper Limit		Q value	df	p
Primary	8	.89	.27	.37	1.41	.00			
Secondary	28	.71	.07	.56	.86	.00			
Tertiary	11	.64	.14	.36	.92	.00			
							.66	2	.71

\*p < .05

# Évaluation par les pairs ?

Revue systématique (Double et al., 2020)

**QR :** (1) Quel est l'effet d'intervention d'évaluation par les pairs sur les performances académiques des élèves comparativement aux performances académiques observées chez les élèves des groupes contrôles ? ; (2) Quelles caractéristiques sont susceptibles de modérer l'efficacité de l'intervention ?

**Méthode :** recherche dans 3 bases de données bibliographiques + Google Scholar → 4072 références  
recherche ancestrale : 37 références  
sélection de 55 références qui rencontrent les CI (Exp. ; Mesure cognitive ; format)

## Résultats :

140 ampleurs de l'effet ont été extraites des 55 études incluses. L'effet moyen calculé est de 0.31 (IC 95% [0.18 ; 0.44]). En d'autres mots, comparés aux étudiants qui n'ont pas participé à l'évaluation par les pairs, les étudiants qui y ont participé ont démontré une augmentation de 31 % d'écart-type de leurs performances académiques.

L'hétérogénéité des effets observés ( $I^2 = 81.08 \%$ ) est importante.

Table 3 Results of the sub-groups analysis

	<i>N</i>	<i>k</i>	<i>g</i>	SE	<i>f</i> <sup>2</sup>	<i>p</i>
Publication type						
Dissertation	8	14	0.21	0.13	64.65%	0.138
Journal	43	123	0.31	0.07	83.23%	< .001
Conference/report	2	3	0.82	0.22	9.08%	0.168
Education level						
Primary school	11	35	0.41	0.12	68.36%	0.006
Secondary	13	22	0.44	0.1	69.70%	0.001
Tertiary	29	83	0.21	0.10	85.17%	0.043
Comparison group						
Teacher assessment	31	65	0.27	0.09	83.82%	0.007
No assessment	23	59	0.31	0.1	78.02%	0.004
Self-assessment	10	16	0.23	0.17	74.57%	0.209
Written						
Yes	36	81	0.35	0.08	84.04%	< .001
No	20	60	0.2	0.08	68.96%	0.014
Dialog						
Yes	19	49	0.21	0.09	70.74%	0.034
No	36	92	0.35	0.08	84.12%	< .001
Grading						
Yes	37	95	0.37	0.07	83.48%	< .001
No	18	46	0.17	0.11	72.60%	0.138
Freeform						
Yes	9	29	0.42	0.16	68.68%	0.03
No	45	112	0.29	0.07	82.28%	< .001

Online							
Yes	22	39	0.38	0.12	83.46%	0.003	
No	33	102	0.24	0.08	80.18%	0.004	
Anonymous							
Yes	23	58	0.27	0.11	82.73%	0.019	
No	29	77	0.25	0.08	70.97%	0.004	
Frequency							
Multiple	34	98	0.37	0.07	81.28%	< .001	
Single	21	43	0.2	0.11	80.69%	0.103	
Transfer							
Far	18	26	0.2	0.13	89.45%	0.124	
Near	23	64	0.42	0.08	72.93%	< .001	
None	23	51	0.29	0.11	84.19%	0.017	
Allocation							
Classroom	41	107	0.31	0.07	78.97%	< .001	
Individual	11	31	0.21	0.13	68.59%	0.14	

$N$  = Number of studies,  $k$  = number of effects,  $g$  = Hedge's  $g$ , SE = standard error in the effect size,  $I^2$  = heterogeneity within the group,  $p = p$  value

# Évaluation par les pairs ?

Revue systématique (Li et al., 2020)

**QR :** (1) Quel est l'effet générale de l'évaluation par les pairs sur l'apprentissage ? ; (2) Quels facteurs sont susceptibles d'influencer cet effet ?

**Méthode :** recherche dans 4 bases de données bibliographiques → 350 références  
sélection de 58 références qui rencontrent les CI (Exp. ; Mesure cognitive ; ES inf.)

## **Résultats :**

134 ampleurs de l'effet ont été extraites des 58 études incluses. L'effet moyen calculé est de 0.291 (IC 95% [0.194 ; 0.388]). En d'autres mots, comparés aux étudiants qui n'ont pas participé à l'évaluation par les pairs, les étudiants qui y ont participé ont démontré une augmentation de 29.1 % d'écart-type de leurs performances générales.

L'hétérogénéité des effets observés ( $Q = 628.095$ ) est significative ( $df = 133, p < 0.001$ ).

Table 1. Variables, frequencies and effect size comparison.

Category	Variable	Category	No. of effect sizes	Mean effect size	SE	95% CI	R <sup>2</sup>
Methodology	Research design	Quasi-experimental	98	.320***	.058	[.206, .434]	0%
		Experimental	36	.216*	.094	[.033, .400]	
	Control condition	Peer assessment vs. no assessment	64	.330***	.071	[.190, .469]	0%
		Peer assessment vs. teacher assessment	50	.260**	.082	[.100, .421]	
		Peer assessment vs. self-assessment	20	.239	.132	[-.020, .499]	
Peer assessment setting	Grade level	Higher education	102	.331***	.056	[.221, .441]	0%
		K-12	31	.150	.105	[-.056, .356]	
	Subject area	Social science and arts	88	.284***	.062	[.163, .406]	0%
		Science and engineering	34	.345**	.098	[.154, .537]	
		Medical and clinical	12	.197	.156	[-.108, .503]	
	Task rated	Essay writing	66	.302***	.072	[.162, .443]	0%
Project, exam, or other	68	.281***	.069	[.147, .415]			
Peer assessment procedure	Assessment mode	Paper-based	100	.237***	.056	[.127, .348]	2.97%
		Computer-mediated	34	.452***	.097	[.261, .643]	
	Rating format	Only scores	23	.374**	.116	[.147, .602]	.27%
		Only comments	51	.176*	.082	[.015, .337]	
	Feedback mode for comments	Both scores and comments	60	.349***	.073	[.206, .491]	
		Written feedback	71	.256***	.065	[.129, .382]	0%
		Oral feedback	21	.205	.120	[-.030, .440]	
	Rater training	Both	19	.415**	.126	[.168, .663]	
		Peer raters did not receive training	37	.017	.090	[-.160, .194]	9.18%
	Peer raters received training	97	.396***	.056	[.286, .505]		
	Rating criteria	Without explicit rating criteria	16	.136	.148	[-.154, .425]	.17%
		With explicit rating criteria	118	.311***	.052	[.208, .413]	
	Frequency	One peer assessment session	53	.206**	.078	[.053, .359]	1.33%
		More than one peer assessment session	81	.347***	.063	[.223, .470]	
	Requirement	Peer assessment is compulsory	105	.293***	.056	[.183, .404]	0%
Peer assessment is voluntary		29	.284**	.105	[.079, .489]		
Assessors and assessees	Reciprocity	Assessor only	20	.161	.124	[-.082, .403]	0%
		Assessee only	6	.197	.233	[-.260, .654]	
		Both assessor and assessee	108	.323***	.055	[.214, .431]	
	Number of assessors per assignment	One assessor per assignment	68	.321***	.070	[.185, .457]	0%
		More than one assessor per assignment	66	.261***	.070	[.122, .399]	
	Matching of assessors and assessees	Raters and assessees are matched at random	58	.337***	.075	[.189, .485]	0%
		Raters and assessees are not matched at random	76	.256***	.066	[.128, .385]	
	Anonymity	Anonymous rating	45	.383***	.086	[.215, .550]	0%
		Non-anonymous rating	89	.246***	.060	[.128, .364]	

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ . $p$  values for the between-category effect size comparisons for each variable were all higher than .05 with the exception of rater training ( $p = .0004$ ).

# Évaluation par les pairs ?

Revue systématique (Li et al., 2021)

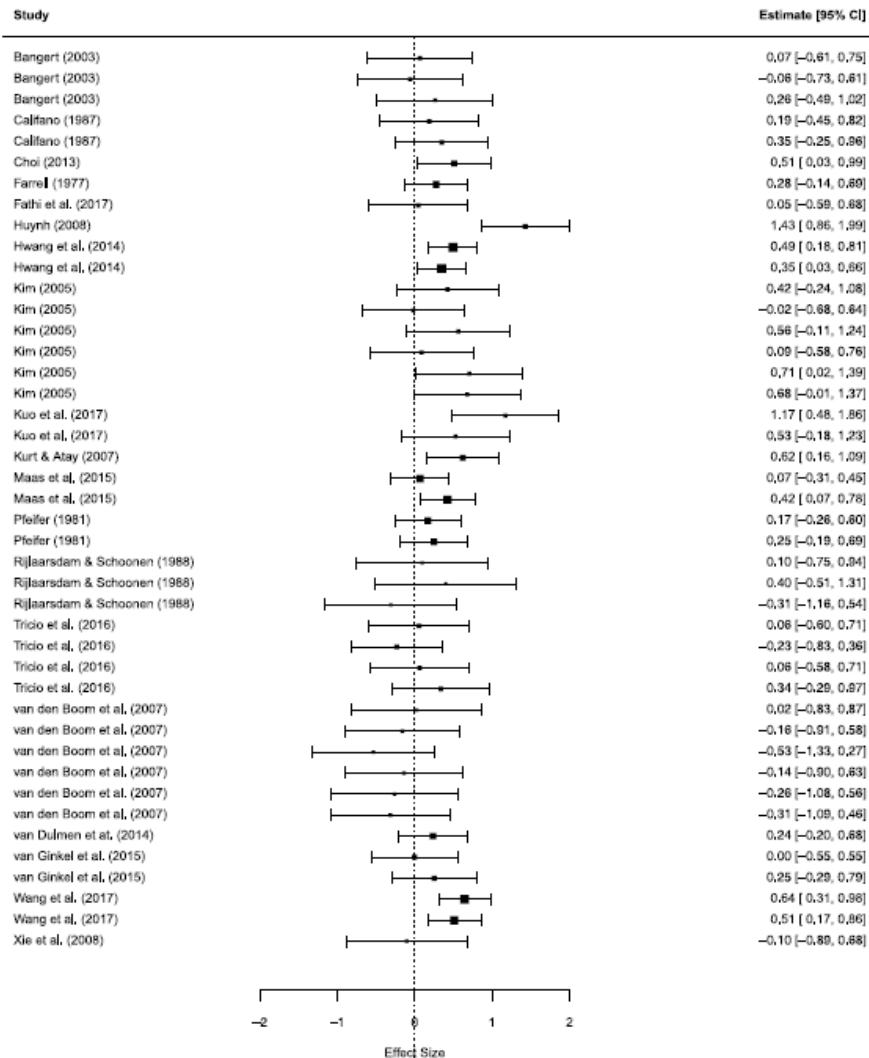
**QR :** (1) Quel est l'effet générale de l'évaluation par les pairs sur les variable non-cognitives ? ; (2) Quels facteurs sont susceptibles d'influencer cet effet ?

**Méthode :** recherche dans 4 bases de données bibliographiques  
sélection de 19 références qui rencontrent les CI (Exp. ; Mesure non-cognitive ; ES inf.)

## **Résultats :**

43 ampleurs de l'effet ont été extraites des 19 études incluses. L'effet moyen calculé est de 0.289 (IC 95% [0.134 ; 0.444]). En d'autres mots, comparés aux étudiants qui n'ont pas participé à l'évaluation par les pairs, les étudiants qui y ont participé ont démontré une augmentation de 28.9% d'écart-type de leurs résultats à des tests non-cognitifs.

L'hétérogénéité des effets observés est importante.



**Table 2.** Variable frequency and meta-regression results for research design and peer assessment setting.

Category	Variable	Meta-regression predictor	Frequency	Intercept			Predictor		
				Value	SE	P value	Regression coefficient	SE	P value
Research design	Experimental design	Quasi-experimental (reference group)	32	0.349	0.070	.001			
		Experimental	11				-0.239	0.186	.259
Peer assessment setting	Control condition	Peer assessment vs. no assessment (reference group)	16	0.320	0.102	.042			
		Peer assessment vs. teacher assessment <sup>a</sup>	25				-0.046	0.151	.769
	Outcome variable	Learning strategies (reference group)	28	0.337	0.073	.000			
		Academic mind-sets	15				-0.146	0.124	.267
	Grade level	Higher education (reference group)	33	0.251	0.091	.021			
K-12		10				0.132	0.125	.331	
Subject area	Writing (reference group)	12	0.372	0.120	.021				
	Non-writing	31				-0.118	0.151	.452	

Each variable was entered to the regression one at a time. SE = Standard Error.

<sup>a</sup>If peer assessment vs. teacher assessment was the reference group, the corresponding effect size was 0.274, SE = 0.111,  $p = .036$ .

**Table 3.** Variable frequency and meta-regression results for peer assessment procedure, assessors and assessees.

Category	Variable	Meta-regression predictor	Frequency	Intercept			Predictor		
				Value	SE	P value	Regression coefficient	SE	P value
Peer assessment procedure	Assessment mode	Paper-based (reference group)	22	0.192	0.041	.001			
		Computer-mediated	21				0.214	0.145	.167
	Rating format	Both scores and comments (reference group)	16	0.439	0.059	.000			
		Either comments or scores	27				-0.276	0.112	.030
	Rater training	Peer raters did not receive training (reference group)	21	0.214	0.110	.100			
		Peer raters received training	22				0.145	0.143	.330
	Frequency	One peer assessment session (reference group)	11	0.259	0.124	.189			
	More than one peer assessment session	32				0.035	0.153	.834	
Assessors and assessees	Requirement	Peer assessment is compulsory (reference group)	22	0.407	0.080	.000			
		Peer assessment is voluntary	21				-0.248	0.132	.086
	Matching of assessors and assessees	Assessors and assessees matched at random (reference group)	25	0.417	0.069	.000			
		Assessors and assessees not matched at random	18				-0.305	0.119	.027
	Anonymity	Anonymous rating (reference group)	17	0.405	0.084	.007			
	Non-anonymous rating	26				-0.185	0.126	.176	
Number of assessors per assignment	One assessor per assignment (reference group)	22	0.322	0.078	.006				
	More than one assessor per assignment	21				-0.073	0.148	.633	

Each variable was entered to the regression one at a time. SE = Standard Error.

# Évaluation par les pairs ?

Revue systématique (Li et al., 2016)

**QR :** (1) Quel est le degré d'accord entre l'évaluation par un pair et l'évaluation par l'enseignant ? ;  
(2) Quels facteurs sont susceptibles d'influencer cet accord ?

**Méthode :** recherche dans 4 bases de données bibliographiques → 292 références  
sélection de 69 références qui rencontrent les CI (mesure numérique de l'accord)

## **Résultats :**

269 ampleurs de l'effet ont été extraites des 69 études incluses. L'effet moyen calculé est de 0.69 ( $p < 0.0001$ ). L'évaluation menée par un pair montre, généralement, un degré d'accord modérément élevé avec l'évaluation menée par un enseignant.

L'hétérogénéité des effets observés ( $X^2 = 2471.35$ ) est significative ( $df = 268, p < 0.001$ ).

Table 2. Results of the final model with only significant predictors.

Predictors	Regression coefficient*	Standard error	<i>T</i> ratio	<i>P</i> value
Intercept	.69	.05	14.44	.000
Assessment mode is computer-assisted (vs. paper-based)	-.14	.06	-2.25	.026
Subject area is medical/clinical (vs. social science/arts)	-.35	.11	-3.19	.002
Course is graduate level (vs. undergraduate level)	.18	.07	2.49	.013
Assesseees are groups (vs. individuals)	-.26	.07	-3.56	.000
Assessors and assesseees are not matched at random (vs. matched at random)	-.27	.07	-3.83	.000
Peer assessment is voluntary (vs. compulsory)	.28	.08	3.40	.001
Peer raters are non-anonymous (vs. anonymous)	.15	.06	2.40	.017
Both scores and comments are provided (vs. only scores are provided)	.15	.06	2.57	.011
Peer raters are involved in developing the rating criteria (vs. peer raters are not involved)	.60	.11	5.55	.000

\*The regression coefficients are in the Fisher's Zeta metric.

# Auto-évaluation?

## Revue systématique (Karaman, 2021)

**QR :** (1) Quel est l'effet des interventions d'auto-évaluation sur les performances académiques des élèves ? ; (2) Quelles sont les variables modératrices qui influencent cette efficacité ?

**Méthode :** recherche dans 2 bases de données bibliographiques + Google scholar → 300 références  
sélection de 18 références qui rencontrent les CI (QED, perf., ES)

### **Résultats :**

46 ampleurs de l'effet ont été extraites des 18 études incluses. L'effet moyen calculé est de 0.37 (CI 95% [0.29 ; 0.46] ;  $p < 0.05$ ). En d'autres mots, comparés aux étudiants qui n'ont pas participé à l'évaluation par les pairs, les étudiants qui y ont participé ont démontré une augmentation de 37 % d'écart-type de leurs performances académiques.

L'hétérogénéité des effets observés ( $Q = 132.94$ ) est significative ( $df = 45, p < 0.05$ ).

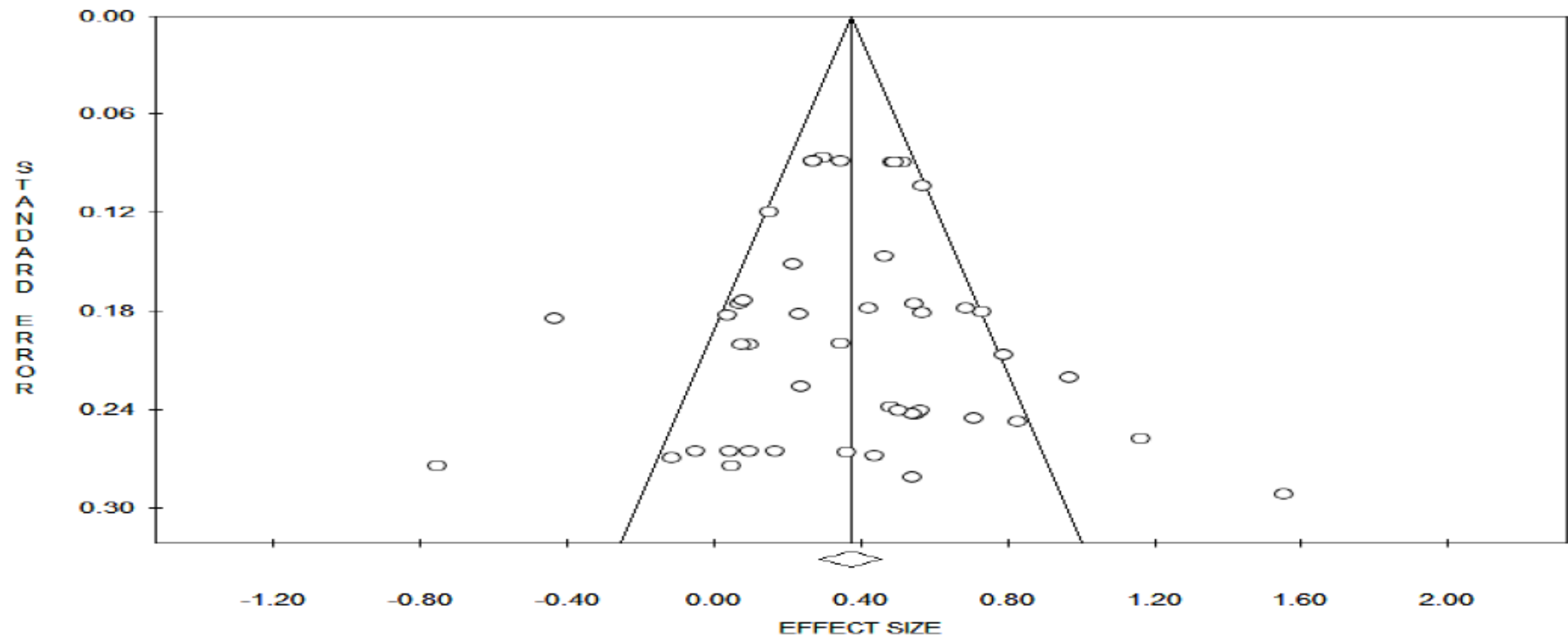


Figure 2. Funnel Plot

Table 2. Results of the Moderator Analysis for Academic Performance

Moderator variable	k	g	95 % CI	p value	SE	N	Heterogeneity		
							Q between	df	p value
<b>Education Level</b>									
Primary	12	.43	[.30, .56]	.00	.07	2215	1.05	2	.59
Secondary	22	.37	[.26, .47]	.00	.05	4737			
Tertiary	12	.28	[-.05, .60]	.09	.16	702			
<b>External feedback</b>									
Yes	27	.28	[.15, .41]	.00	.06	3532	5.86*	1	.01
No	19	.47	[.38, .57]	.00	.05	3726			
<b>Self-assessment criteria</b>									
Not using any assessment tool	28	.36	[.27, .45]	.00	.05	5801	.31	1	.57
Using specific assessment tool (rubrics, scripts etc.)	18	.42	[.23, .61]	.00	.10	1853			
<b>Self-assessment training</b>									
Yes	13	.36	[.20, .51]	.00	.08	3867	.05	1	.81
No	33	.38	[.28, .48]	.00	.05	3787			

\*p&lt;.05; k= number of effects; SE= standard error

# Votre/vos question(s)

Comment la (l'évaluation formative) rendre plus efficace dans nos classes ?

# Votre/vos question(s)

Que pensez-vous du redoublement dans la lutte contre l'échec scolaire ?

Sans redoublement, l'élève aura le sentiment que même en ne travaillant pas ou en obtenant des résultats médiocres, on passe quoiqu'il arrive.

Où est la place des parents dans l'échec scolaire ? Que faire pour lutter contre le milieu socio-culturel ?

- Déterminisme sociale et discontinuité culturelle
- VS. discrimination systémique

Comment rendre efficaces les dispositifs d'aide aux élèves en difficulté ?

Que faire si les difficultés de l'élève nous dépassent ?

Que peut-on faire à l'école maternelle pour permettre aux enfants d'avoir un niveau commun sans aide extérieure ?

Scolarité par niveaux avec des classes ouvertes : utopie ?

# Bibliographie

# Bibliographie

- Anderson, L.W., & Krathwohl, D.R. (2001). *A taxonomy for Learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Addison Wesley Longman
- Astolfi, J.-P. (2020). *Lerreur, un outil pour enseigner* (13<sup>th</sup> ed.). Pratiques et enjeux pédagogiques.
- Autin, F., Batruch, A., & Butera, F. (2019). The function of selection of assesment leads evaluators to artificially create te social class achievement gap. *Journal of Educational Psychology*, 111, 717-735. <https://doi.org/10.1037/edu0000307>
- Baye, A., & Dachez, D. (2022). *A randomized controlled trial on the "covid generation": were students' assessments biased during the health crisis* [Paper presentation]. BAPS 2022, KU Leuven, Leuven, Belgium
- Baye, A., & Monseur, C. (2016). Gender differences in variability and extreme scores in an international contexte. *Large-scale Assessments in Education*, 4, 1-16. <https://doi.org/10.1186/s40536-015-0015-x>
- Boutand, J., & De Lara, M. (2019). *Les biais de l'esprit: comment l'évolution a forgé notre psychologie ?* Odile Jacob
- Coffield, F., Moseley, D., Hall, E., & Ecclestone, K. (2004). *Learning styles and pedagogy in post-16 learning : a systematic review and critical review*. Learning and Skills Research Centre.
- Courtier, P., Gardes, M.-L., Van der Henst, J.-B., Novek, I. A., Crosset, M.-C., Epinat-Duclos, J., Léone, J., & Prado, . (2021). Effects of Montessori Education on the Academic, Cognitive, and Social Development of Disadvantaged Preschoolers: A Randomized Controlled Study in the French Public-School System. *Child Development*, 92(5), 2069-2088. <https://doi.org/10.1111/cdev.13575>

# Bibliographie

- Crahay, M. (2019) *Peut-on lutter contre l'échec scolaire ?* (4th ed.). De Boeck supérieur
- Double, K. S., McGrane, J. A., & Hopfenbeck, T. N. (2020). The impact of peer assessment on academic performance: a meta-analysis of control group studies. *Educational Psychology Review*, 32, 481-509. <https://doi.org/10.1007/s10648-019-09510-3>
- Durand, T. C. (2019). *Quand est-ce qu'on biaise ?*. Paris, France : HumenSciences (société)
- Kahneman, D. (2012). *Système 1, système 2 : les deux vitesses de la pensée*. Flammarion
- Kahneman, D. Sibony, O., & Sunstein, C. R. (2021). *Noise*. Odile Jacob
- Karaman, P. (2021). The effect of formative assessment practices on student learning: a meta-analysis study. *International Journal of Assessment Tools in Education*, 8(4), 801-817. <https://doi.org/10.21449/ijate.870300>
- Karaman, P. (2021). The impact of self-assessment on academic performance: a meta-analysis study. *International Journal of Research in Education and Science*, 7(4), 1151-1166. <https://doi.org/10.46328/ijres.2344>
- Lafontaine, D., & Monseur, C. (2009). Les évaluations des performances en mathématiques sont-elles influencées par le sexe de l'élève ? *Mesure et Evaluation en Education*, 32 (2), 1-106. <https://doi.org/10.7202/1024955ar>
- Lee, H., Chung, H. Q., Zhang, Y., Abedi, J., & Warschauer, M. (2020). The effectiveness and features of formative assessment in US K-12 education: a systematic review. *Applied Measurement in Education*, 33(2), 124-140. <https://doi.org/10.1080/02602938.2018.1545896>
- Li, H., Bialo, J. A., Xiong, Y., Hunter, C. V., & Guo, X. (2021). The effect of peer assessment on non-cognitive outcomes: a meta-analysis. *Applied Measurement in Education*, 34(3), 179-203. <https://doi.org/10.1080/08957347.2021.1933980>

# Bibliographie

- Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment & Evaluation in Higher Education*, 45(2), 193-211. <https://doi.org/10.1080/02602938.2019.1620679>
- Li, H., Xiong, Y., Zang, X., Kornhaber, M. L., Lyu, Y., Sun Chung, K., & Huen, H. K. (2016). Peer assessment in the digital age: a meta-analysis comparing peer and teacher ratings. *Assessment & Evaluation in Higher Education*, 41(2), 245-264. <https://doi.org/10.1080/02602938.2014.999746>
- Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: the effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37(11), 2098-2109. <https://doi.org/10.1037/0022-3514.37.11.2098>
- Morrissette, J. (2013). Des modes d'interaction au cœur de la mise en œuvre d'une évaluation formative non instrumentée. *Nouveaux Cahiers de la Recherche en Education*, 16, 88-111. <https://doi.org/10.7202/1029143arv>
- Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles: concepts and evidence. *Psychological Science in the Public Interest*, 9(3), 105-119. <https://doi.org/10.1111/j.1539-6053.2009.01038.x>
- Rangvid, B. S. (2015). Systematic differences across evaluation schemes and educational choice. *Economics of Education Review*, 48, 41-55. <http://dx.doi.org/10.1016/j.econedurev.2015.05.003>
- Sander, E., Gros, H., Gvozdic, K., & Scheibling-Sève, C. (2018). *Les neurosciences en éducation*. Retz (coll. Mythes et Réalités).

# Bibliographie

- Sprietsma, M. (2013). Discrimination in grading: Experimental evidence from primary school teachers. *Empirical Economics*, 45, 523–538. <http://dx.doi.org/10.1007/s00181-012-0609-x>
- Sunstein, C. R., Bobadilla-Suarez, S., Lazzaro, S. C., & Sharot, T. (2017). How People Update Beliefs about Climate Change: Good News and Bad News. *Cornell Law Review*, 102(6), 1431-1443. Retrieved from: <https://scholarship.law.cornell.edu/clr/vol102/iss6/1/>